

McDowell's Realism

Paul Broadbent

a thesis submitted for the degree of
Doctor of Philosophy
at the University of Otago, Dunedin,
New Zealand.

July 2013

Abstract

John McDowell's work presents stimulating arguments on a broad variety of issues. In this thesis I shall look at three: his response to Dummett's manifestation argument (the argument that our understanding of truth must be evidence-constrained because we cannot manifest grasp of verification transcendent truth-conditions); his response to Wittgenstein's rule-following problem (the problem of how meaning is possible given that there don't seem to be any facts which constitute having a meaning in mind); and his argument for moral realism in response to Mackie and Blackburn.

In each of these debates I shall focus on McDowell's central papers and the work of the philosophers which he engages with. Because of the dense and elusive nature of much of McDowell's writing I shall engage in a constructive exegesis of his arguments in order to evaluate them. This work often reveals what seem to be troubling gaps in the justification for his claims, but I am usually able to draw on his other work to provide the required justification.

I conclude that McDowell's response to Dummett's manifestation argument is a success: he is right to claim that we are able to manifest our understanding of sentences simply by saying them to speakers of our language. I explain how McDowell is able to provide justification for the controversial premises in this argument by using his argument that theories of meaning must be modest, and shall use his arguments for fallibilism to save his argument from a fatal ambiguity. However, this reveals a weakness in McDowell's understanding of mathematics, and I argue that his is wrong to endorse mathematical anti-realism.

Next I shall turn to McDowell's response to Wittgenstein's rule-following problem. I start by explaining why Kripke's and Wright's responses to the problem fail, and use that to describe the shape a successful response to the problem must take. I shall argue that McDowell provides a successful response to the rule-following problem, but only if we are happy with his epistemic fallibilism. I attempt to make McDowell's quietist response to the problem more palatable by looking at it in the context of his overall view of the world and our experience of it. And I shall conclude by arguing that in order for McDowell to make sense of our making certain mistakes about our inner states he needs to make some changes to his account of the inner.

Finally I shall look at McDowell's moral realism. I shall explain his response to

Mackie's arguments for queerness which draws a parallel between values and secondary qualities, and to do so I shall develop his no-priority view of values. I shall then turn to Blackburn's quasi-realism and argue that McDowell's attacks of Blackburn, the contaminated response and disentangling arguments, all fail to demonstrate that Blackburn's starting point is incoherent. However, I also defend McDowell from Blackburn's criticism, focusing on his epistemology of susceptibility to reasons, and by drawing parallels with the epistemology of mathematics, using my conclusions from the first chapter.

Acknowledgements

A great many people have provided me with essential help and support in writing this thesis. Primarily I would like to express my gratitude for the support, guidance, and insight provided by my supervisor Prof. Alex Miller. I would also like thank Dr. Charles Pigden, my secondary supervisory, for providing me with some essential feedback towards the end of this project. There are many others I could thank for providing me with support and feedback during this process, but to avoid leaving anyone out I shall simply name those who have made the most radical impact on the content of this thesis, through feedback on chapter drafts and stimulating discussion: Christopher Devereux, Ben Bessey, Toby Betenson, James Darcy, and Daniel Wee. I would also like to thank the post-graduate communities and staff at both Birmingham and Otago for being so supportive.

Additional thanks to the Universities of Birmingham and Otago for providing the funding which has made this thesis possible. And further special thanks go to Prof. Alex Miller for providing funding for my fees once I moved to the University of Otago, out of his own research grant money.

And finally, particular gratitude for Verity Carver, for proofreading the entire thesis in such a short amount of time and for her love and support.

Contents

Introduction	1
1 The manifestation argument	8
1.1 The three realist commitments	9
1.1.1 The Wittgensteinian conception of linguistic understanding	10
1.1.2 The truth-conditional conception of understanding	14
1.1.3 Evidence-transcendent truth	16
1.2 The significance of semantic realism	19
1.3 The acquisition argument	22
1.3.1 The realist's response	23
1.4 The manifestation argument	24
1.4.1 Core and neighbourhood abilities	25
1.4.2 The argument	27
1.4.3 The manifestation challenge	28
1.4.4 The anti-realist solution	29
1.5 McDowell's response: the argument from the truth-assertion platitudes	30
1.6 Solution to problem 1: modesty	34
1.7 Problem 2: epistemically constrained truth	37
1.7.1 Figuring in speech	40
1.7.2 McDowell's argument for anti-realism about mathematics	41
1.7.3 Wright's epistemological mistake and McDowell's fallibilism	43
1.7.4 The link between fallibilism and realism	46
1.7.5 The solution to problem 2	47
1.8 Problem 3: mathematical realism	48
1.8.1 The problem with c	49
1.8.2 The significance of the falsity of c	51
1.9 Conclusion	52
2 Rule-following part 1: The problem of rule-following	54
2.1 Kripke's sceptical challenge	55
2.1.1 Dispositions	58
2.1.2 Internal states	59
2.1.3 Kripke's sceptical solution	62
2.2 Responses to Kripke	65
2.2.1 McDowell	65
2.2.2 Wright	67
2.2.3 Summary	69
2.3 Wright's communitarian solution	69

2.3.1	Wright’s argument	71
2.3.2	Wright’s solution	73
2.4	Criticisms of Wright’s communitarian view	74
2.4.1	McDowell	74
2.4.2	Boghossian	76
2.5	Wright’s judgement-dependent account	78
2.5.1	Wright’s attack of Kripke’s inferential assumption	78
2.5.2	Judgement-dependence	80
2.5.3	The judgement-dependent account of intentions	83
2.5.4	The judgement-dependent account of meaning	87
2.6	Evaluation of Wright’s judgement-dependent account	88
3	Rule-following part 2: McDowell’s solution	95
3.1	McDowell’s diagnosis of Wright’s mistake	98
3.2	The master thesis	99
3.3	Conceptually structured experience	103
3.3.1	The meaning of “conceptually structured”	108
3.4	Conceptually structured inner awareness	109
3.5	Knowledge of meaning and epistemic fallibilism	111
3.5.1	Interpretations and disjunctivism	113
3.6	Quietism	115
3.7	McDowell’s solution as involving <i>sui generis</i> mental states	119
3.8	The role of the community	120
3.9	The seems right / is right distinction	122
3.10	Conclusion	128
4	McDowell’s moral realism part 1: Mackie’s error theory	131
4.1	Mackie’s error theory	132
4.1.1	Mackie’s arguments for error theory	135
4.2	McDowell’s no-priority view	141
4.2.1	The two notions of objectivity	143
4.2.2	Williams’ absolute conception	146
4.2.3	McDowell’s siblings analogy	152
4.3	McDowell’s response to Mackie’s arguments against moral realism	153
4.3.1	The argument from relativity	153
4.3.2	The argument from queerness	158
4.4	Summary	164
5	McDowell’s moral realism part 2: Blackburn’s quasi-realism	166
5.1	Blackburn’s quasi-realism	167
5.1.1	Blackburn’s arguments for non-cognitivist projectivism	168
5.2	Quasi-realism	175
5.2.1	The Frege-Geach problem	175
5.2.2	Earning truth	177
5.2.3	The problem of mind dependence	179
5.2.4	The conflict between quasi-realism and non-cognitivism	180
5.3	Explanatory tests	183
5.3.1	The significance of the quasi-realist starting point	183

5.3.2	The two different explanatory tests	185
5.4	The first explanatory test: McDowell's contaminated response and disentangling arguments	187
5.4.1	The disentangling argument	188
5.4.2	The contaminated response argument	196
5.4.3	Summary	202
5.5	The second explanatory test: the explanatory failure of moral facts	204
5.5.1	Blackburn's challenge	205
5.6	The epistemology of susceptibility to reasons	206
5.6.1	Earning truth in ethics	208
5.6.2	The parallel with mathematics	209
5.7	Conclusion	212
	Conclusion	214
	References	221

Introduction

John McDowell is one of the foremost figures in analytic philosophy of the past thirty-five years. His work presents stimulating arguments on a broad variety of issues, such as philosophy of language, metaphysics, ethics, metaethics, epistemology, and the philosophy of mind. Whilst his research is very compelling, his arguments are notoriously difficult to pin down; they are often dense and any single argument tends to draw on his views from all across philosophy. Given the complexity of McDowell's work and word limit placed on a PhD thesis it would be impossible for me to go into sufficient detail on all of McDowell's work, or even all his work relating to realism. And so in order to restrict the focus of this thesis, and to enable me to look at particular areas in greater depth, I shall look only at the topics McDowell engaged with relating to realism in the 1970s and 1980s. Moreover, although the realism / anti-realism disputes have continued to grow since the 1980s, the forms of opposition to realism discussed by McDowell — expressivism, error-theory, and semantic anti-realism — remain the canonical starting points for contemporary discussions of realism and its opponents.

The key debates relating to realism which McDowell entered into during that period are those of the manifestation argument, meaning, and value. When looking at the manifestation argument, I shall focus on his response to Dummett and Wright; when looking at meaning I shall focus on his solution to the rule-following problem by contrasting it with Kripke and Wright's solutions; and finally, I shall turn to look at his development of a moral realism in response to Mackie and Blackburn. To get the most charitable reading of McDowell's view I shall sometimes draw on McDowell's work outside the period of the 1970s and 1980s, but this will usually be used to flesh out, or subtly improve, minor points in these arguments, and shall not be the primary focus of this thesis.

Because of the elusive nature of much of McDowell's writing, I shall be engaging in constructive exegesis of his work in order to evaluate the success of his arguments. I shall look at McDowell's core papers on each of the three topics, explain the background debate that he is responding to, and make his arguments explicit. By drawing out his arguments I often reveal what seem to be glaring holes and unjustified assumptions, but I typically conclude that these assumptions can be justified by turning to other areas of his philosophical research, sometimes even relying on work which wasn't published when the original papers were. Although this work typically leaves McDowell's arguments in a

strong position, I also reveal conflicts between aspects of his view, and further work that needs to be done.

I shall present a brief overview of my chapters on the three key topics I focus on, before turning to look at the underlying methodology.

Chapter 1: the manifestation argument

The manifestation argument attempts to show that if we endorse Wittgenstein's claims about the necessary publicity of language then we must reject realism. I focus on McDowell's response to this argument in "Anti-realism and the Epistemology of Understanding" (McDowell, 1981/1998b), but by making the argument explicit I reveal that two of the key premises stand unjustified. I turn to two of McDowell's later papers to show how we can justify these premises, but this leads to a conflict with McDowell's view of mathematics.

According to the Wittgensteinian conception of linguistic understanding the facts about how we understand sentences are constituted by the abilities that we are able to publicly manifest. Wright, in his refinement of Dummett's original argument, argues that we cannot publicly manifest our ability to understand what it is for a sentence to be undetectably true or false, and so our understanding of truth must be evidence-constrained. This is problematic for realism, because realism requires us to be able to make sense of the idea that there may be truths that potentially outrun our capacities to gather evidence; if we cannot make sense of this idea then we do not have a conception of truth which goes beyond what we can know — reality seems to be constrained by our access to it. And, as Miller (2003b) argues, if our sentences cannot have truth-values beyond those for which we can potentially have evidence for, then how are we able to make sense of realism? If our understanding of our sentences is evidence constrained, then so is reality as we understand it. And so we cannot make sense of realism.

McDowell responds to this argument by arguing that, contrary to Wright and Dummett, we don't need to be able to test whether a sentence's truth-conditions obtain to be able to manifest our understanding of it; instead we can manifest our understanding of sentences simply by asserting them. McDowell's argument for this claim comes from a particularly dense passage in "Anti-realism and the Epistemology of Understanding" (McDowell, 1981/1998b); I shall unpack the argument, and in doing so reveal three potential problems.

The first problem I identify is that the first premise of the argument is highly controversial, despite McDowell's claim that it is a mere platitude about assertion. The key premise I identify is that "Knowledge of what a sentence can be used to assert is knowledge which can be directly manifested ... by using the sentence ... to assert precisely that." (McDowell, 1981/1998b, p. 321). However, I respond on McDowell's behalf by arguing that we can defend this premise by drawing on his work in the paper "In Defence

of Modesty” (McDowell, 1987/1998e).

The second problem with McDowell’s argument is that the second premise is ambiguous, and this ambiguity stops McDowell’s argument in its tracks. However, like the problem with the first premise, I argue that we can draw on McDowell’s other work — this time his argument for epistemological fallibilism in the paper “On “The Reality of the Past” ” (McDowell, 1978/1998k) — to solve the problem. This leads me to the conclusion that McDowell’s argument for realism only works in domains in which a fallibilist account of epistemology is appropriate, and so reveals an additional requirement for a successful response to the manifestation argument.

Finally, I argue that a consequence of my response to the second problem is that we should be realists about mathematics. This is problematic because this conclusion follows solely from a combination of McDowell’s arguments, and yet directly conflicts with McDowell’s rejection of mathematical realism in “Mathematical Platonism and Dummettian Anti-Realism” (McDowell, 1989/1998g). I explain why we should be realists about mathematics, and suggest where McDowell’s argument for anti-realism about mathematics goes wrong.

Chapters 2 & 3: rule-following

Kripke’s *Wittgenstein on Rules and Private Language* (Kripke, 1982) famously interprets Wittgenstein’s rule-following arguments in the *Philosophical Investigations* (Wittgenstein, 1953/2001) as attempting to demonstrate that there are no facts that constitute having a particular meaning in mind. An essential aspect of meaning is that it is normative — there are right and wrong ways to use language — but Kripke argues that there is no fact which constitutes our having a particular meaning in mind that is able to provide the required normative constraints. This leaves us with the question of how meaning is possible. In the first of these two chapters, I shall look at Kripke’s original argument, three attempted solutions to the problem (Kripke’s, and two from Wright), and argue that all three fail.

However, these three attempted solutions are useful in that they reveal four requirements upon any successful response to the problem of meaning. In the second of these chapters, I turn to McDowell’s response to the rule-following problem, bringing together the arguments he makes across the papers “Wittgenstein on Following a Rule” (McDowell, 1984/1998o), “One Strand in the Private Language Argument” (McDowell, 1989/1998j), “Intentionality and Interiority in Wittgenstein” (McDowell, 1991/1998f), and “Meaning and Intentionality in Wittgenstein’s Later Philosophy” (McDowell, 1993/1998h) and look at how he satisfies the four requirements. Like Wright’s second approach, McDowell endorses a non-reductive account of meaning — he thinks that Kripke’s original argument is guilty of assuming that knowledge of meaning must be inferential, and as Wright demon-

strates, this is a very questionable assumption. Whereas Wright attempts to explain the possibility of meaning by claiming that what we mean is constituted by our best judgements, McDowell instead develops a quietist response to the problem. This is the attempt to remind us of common sense facts to show that we shouldn't be worried about how it is possible for us to have non-inferential knowledge of what we mean.

I attempt to make McDowell's response more palatable by drawing parallels with his view of the world and our experience of it, which he developed in *Mind and World* (McDowell, 1994/1996). And I shall explain the role McDowell's epistemological views play in justifying this view, highlighting the significance of his endorsement of direct realism, fallibilism, and epistemic internalism.

I shall also look at the role the community plays in McDowell's response to rule-following. McDowell makes sense of the possibility of mistakes in our use of language by claiming that all of language is necessarily thinkable from second or third person points of view. Despite first appearances, this is compatible with the idea that a person born in isolation is capable of grasping meanings and following rules.

However, I shall argue that McDowell's response to the need for a seems right / is right distinction conflicts with what he says about inner states, and that his account of the inner seems to make it impossible for us to make certain types of mistakes about them. I argue that for McDowell's response to the rule-following problem to succeed he needs to drop his claim that inner objects "have no existence independently of [the awareness of "inner experience"]" (McDowell, 1994/1996, p. 21), but this leaves him with the requirement of developing a new account of the distinction between the inner and the outer.

Chapters 4 & 5: McDowell's moral realism

In the final two chapters I turn to McDowell's moral realism, looking at his view by seeing how he responds to Mackie's arguments against moral realism, and his criticisms of Blackburn's quasi-realism. In *Ethics: Inventing Right and Wrong* Mackie (1977) argues that our moral judgements express beliefs about the world, but that all atomic positive moral judgements are false because there are no moral facts. He argues for this conclusion by claiming that we can best explain our moral experience as being a result of our projecting moral properties onto the world, rather than our detection of moral properties. This is because moral beliefs seem to be relative to social groups, and their reality would be both metaphysically and epistemologically queer.

McDowell responds to Mackie's arguments in the papers "Aesthetic Value, Objectivity, and the Fabric of the World" (McDowell, 1983/1998a), and "Values and Secondary Qualities" (McDowell, 1985/1998m). But I argue that the best way to understand what unifies his responses in these papers is by focusing on the "no-priority" view, which only

becomes clearly identified in the paper “Projection and Truth in Ethics” (McDowell, 1987/19981) (a paper that is actually a response to Blackburn). I shall draw on all three of these papers to develop my account of McDowell’s no-priority view and then use it to explain how McDowell uses it respond to Mackie’s arguments against realism.

Like Mackie, Blackburn argues that we should start off by understanding our moral experience as being a result of our projecting moral properties onto the world. However, he differs from Mackie in that he thinks we should start off by understanding moral judgements to be primarily expressing sentiments rather than beliefs — he is a non-cognitivist about values.¹ What’s surprising about Blackburn is his quasi-realist project, which is the attempt to go from a non-cognitivist projectivist starting point but from there earn our right to the notion of truth in ethics, and explain how our moral judgements can be true or false.

Blackburn’s arguments for how we can earn truth in ethics parallel some aspects of McDowell’s account, which we’ll see when looking at how McDowell responds to Mackie’s criticisms of the epistemology of moral realism. However, Blackburn attempts to earn our right to the notion of moral truth from a more austere metaphysical starting point that avoids presuming that there are moral properties. McDowell responds to Blackburn by arguing that the only available explanation of our moral judgements and experience involves genuine moral properties right from the start — we cannot begin to explain moral judgements without presuming that there are moral properties. He argues that Blackburn’s metaphysical austere starting point is incoherent, and so there is no need for the quasi-realist’s project.

I express the key conflict between Blackburn and McDowell in terms of their responses to two different explanatory tests: whether the explainer can consistently deny the reality of values; and whether value facts pull their weight in an explanation. I’ll start by looking at McDowell’s two arguments (the disentangling and contaminated response arguments) which attempt to show that Blackburn’s metaphysically austere starting point fails the first of these explanatory tests. However, I shall conclude that both of McDowell’s arguments are failures because they depend upon premises which Blackburn can simply reject. In response, I develop a new version of the contaminated response argument which relies on a less problematic premise, but is still likely to be rejected by Blackburn.

Finally, I turn to the second explanatory test. Blackburn argues that McDowell fails this test — he can’t explain how values pull their weight because he can’t justify his right to endorse the conditional “if it hadn’t been the case that p , I would not be committed to p ” (Blackburn, 1985/1993b, p. 161). I explain McDowell’s response to this criticism by explaining his account of the epistemology of susceptibility to reasons. This view may

¹Blackburn actually rejects the label of his position as “non-cognitivist”. In this chapter I shall look at this controversy and explain why I sometimes refer to aspects of his position as non-cognitivist.

initially appear a little spooky, so I shall attempt to dispel this worry by drawing a parallel with the epistemology of mathematics which makes use of the view of mathematics that I developed at the end of the first chapter of the thesis.

Methodological approach

As I said above, my strategy in response to the elusive nature of McDowell's work has been to focus on the central papers of these three debates, both by McDowell and the other philosophers with whom he engages. I engage in a constructive exegesis of McDowell's work, often drawing on his other work to supplement his arguments.

Although I have decided to focus in-depth on the core original papers in these debates, there are many different ways I could have approached this project. For example, I could have instead looked at McDowell's position by contrasting it with other philosophers who hold similar views. People including Putnam, Rorty, Brandom, Wiggins, Sellars, Wittgenstein, Hegel, Kant, and Aristotle all have aspects of their work which have strong parallels with McDowell's philosophical position. It would be an interesting and valuable project to look at McDowell in the context of his contemporaries whose views have illuminating connections and contrasts with his own. Another related project here would be to look at the influence historical philosophers have had on McDowell, since he often develops his arguments in the context of these great historical figures.

Another limitation of my approach is that it has prevented me from taking an exhaustive look at the secondary literature on McDowell. I have instead chosen to focus in as much depth as possible on a few core papers written by McDowell and to look at the work with which he himself engages to provide the best context for understanding these papers.

Also, given the breadth of McDowell's work there are many important topics I haven't had space to explore. A more complete, longer, evaluation of his work would also spend considerable time exploring his philosophical methodology: McDowell's quietism sets him apart from many contemporary philosophers and plays a key role in his response to many of these debates. In the conclusion I shall briefly draw together the places in which McDowell's quietism plays a major role, and use it to argue that there is not a unified notion of realism running through his position. McDowell sees his role as a philosopher not to be that of providing explanations of how there can be certain types of facts, or how we are able to know about them. Instead, he aims to correct the mistakes which make such explanations seem to be required.

Although McDowell cannot be said to endorsing a unified notion of realism in each of the three topics I look at, there could at least be said to be a unified philosophical approach. One result of my research is that what turns out to be playing a key role in each of these three debates is his epistemology. This would seem to indicate that

epistemology, and not philosophy of language, is the cornerstone of his work; this would be a very worthwhile focus for any future work. Unfortunately, in these pages I have been forced to look at McDowell's epistemology only briefly, and my strategy has been to simply present the arguments for his epistemological claims to show how they relate to these debates, but not to evaluate them in any depth.

Chapter 1

The manifestation argument

The manifestation argument attempts to show that if we endorse the Wittgensteinian conception of linguistic understanding — that linguistic understanding is constituted by our publicly manifestable abilities — then we must reject realism. The key claim here is that we can't manifest our ability to grasp potentially evidence-transcendent truth-conditions, and so our sentences can't have potentially evidence-transcendent truth-conditions. But, for realism to be true our sentences must have potentially evidence-transcendent truth-conditions: they must be capable of being true or false even in cases in which we don't have evidence for their truth-value (or a finite method guaranteed to detect their truth-value). This is because realism requires us to be able to make sense of the idea that there may be truths that potentially outrun our capacities to gather evidence; if we cannot make sense of this idea then we do not have a conception of truth which goes beyond what we can know — reality seems to be constrained by our access to it. And if we don't understand sentences as having truth-values beyond those we can potentially have evidence for then our understanding of reality will be equally constrained (I shall explain the argument for this, using Miller's paper "The Significance of Semantic Realism" (Miller, 2003b) in §1.2).

The manifestation argument was originally developed by Michael Dummett,¹ later finessed by Crispin Wright,² and aims to show that realism is incompatible with the necessary publicity of language. In "Anti-realism and the Epistemology of Understanding" (McDowell, 1981/1998b), John McDowell responds to Dummett, arguing that we can demonstrate realism by simply referring to certain platitudes relating the notions of assertion and truth. However, Wright's criticism of McDowell in the introduction to *Realism*,

¹Dummett's classic papers on the manifestation argument include: "Truth" (Dummett, 1959), "What is a Theory of Meaning (I)" (Dummett, 1976), "Realism" (Dummett, 1963/1978b), "The Reality of the Past" (Dummett, 1969), and "The Philosophical Basis of Intuitionistic Logic" (Dummett, 1973/1978a). For a good overview of the debate see also Hale's "Realism and its Oppositions" (Hale, 1997).

²For examples of Wright's development see: "Realism, Truth-value Links, Other Minds and the Past" (Wright, 1980/1993c), "Misconstruals made Manifest" (Wright, 1989/1993b), and the introduction to *Realism, Meaning and Truth* (Wright, 1993a).

Meaning & Truth (Wright, 1993a) reveals one of McDowell's premises to be deeply ambiguous, and that McDowell's use of it amounts to begging the central question. Another problem for McDowell is that in the paper "Mathematical Platonism and Dummettian Anti-Realism" (McDowell, 1989/1998g) he claims that we should be anti-realists about mathematics, but his argument in "Anti-realism and the Epistemology of Understanding" (McDowell, 1981/1998b) appears to support realism generally, across all assertoric discourses that permit the formation of sentences that are not effectively decidable. There seems to be no reason for that argument not to apply to mathematics, since it is an assertoric practice containing sentences that are not effectively decidable. In this chapter I shall look at these two problems. In response to the first I'll argue that McDowell's later paper contains the required resources to respond to Wright's attack, but can only do so by bringing in substantial epistemological machinery that is highly contested and far from platitudinous. In response to the second problem I shall argue that McDowell's view of mathematics results from a naive understanding of actual mathematical practice, and that once this is corrected it becomes clear that he should be a realist about mathematics. But before we get into the details of McDowell's position, it would be useful to start with an explanation of the manifestation argument, and to look at its relation to Dummett's other argument against realism, the acquisition argument.

1.1 The three realist commitments

At its core, the manifestation argument aims to show that realism involves a commitment to three claims which are incompatible. These are:

1. **The Wittgenstein conception of linguistic understanding:** Linguistic understanding consists in a collection of practical abilities, and there is nothing more to our understanding than what can be manifested publicly; "Meaning cannot transcend use" (McDowell, 1981/1998b, p. 344).
2. **The truth-conditional conception of understanding:** Understanding a declarative sentence consists in grasp of its truth-conditions.
3. **Evidence-transcendent truth:** Truth is unconstrained by the availability of evidence; grasp of the truth-conditions of a sentence may require an understanding of how it could be undetectably true or false.

Dummett, Wright, and McDowell all wholeheartedly endorse **1**,³ so the debate revolves around what its truth implies for the others. In Dummett's original formulation of the

³Of course, Wittgenstein's general view of linguistic understanding isn't universally accepted. For example Fodor (see "A Theory of Content I" (Fodor, 1992b) and "A Theory of Content II" (Fodor, 1992a)), and Chomsky (see *Knowledge of Language* (Chomsky, 1986)) would both reject Wittgenstein's

argument, he claimed that we must respond to the incompatibility of these three by rejecting **2**. This means that he accepted the realist’s claim about the nature of truth in **3**, and that we should instead reject the claim that truth plays a core role in the correct theory of meaning. Later refinements developed by Wright in “Realism, Truth-value Links, Other Minds and the Past” (Wright, 1980/1993c) led him to argue that in fact **2** is compatible with **1**, if we reject **3** and instead reform the notion of truth involved to be anti-realist one. Along with this change he was led to conclude that the manifestation argument cannot be seen to be a proof of the falsity of realism — the three commitments are not logically incompatible —, but instead merely demonstrates that the realist has no justification for endorsing their position.⁴

In this section I will go over each of these three commitments, and the justification for them, before going on to explain Wright’s argument against realism.

1.1.1 The Wittgensteinian conception of linguistic understanding

At the core of Wittgenstein’s view is the claim that we should associate meaning with use — we understand a sentence’s meaning if and only if we have the appropriate collection of practical abilities. Most important for the manifestation argument is the claim that the exercise of the practical abilities to use language must be open to view — the abilities must be manifested in a way that is publicly evaluable as correct or incorrect.

Wittgenstein’s argument for associating meaning with use primarily focuses on a criticism of alternative ways of understanding meaning: the mentalistic and dispositional views.⁵

The mentalistic view of meaning is the view that meaning is constituted by something which comes before the mind, such as an image. For example, a supporter of this view might claim that we only understand the word ‘square’ if we have a disposition to conjure up a mental image of a square in the relevant circumstances. When judging whether the

claim that grasp of meaning is fundamentally a publicly manifestable ability. The manifestation argument would be of no threat to any realist who rejected this view of meaning. However, Wright, Dummett and McDowell all accept a general Wittgensteinian view of language, and in this chapter my goal is to evaluate McDowell’s response to the manifestation argument, so I’m going to simply presume the Wittgensteinian picture is correct. (Dummett, Wright, and McDowell would hold that Wittgenstein’s rule-following arguments refute the kind of views which Fodor and Chomsky endorse, although of course these arguments are also deeply controversial. I shall go into the rule-following arguments over the next two chapters, although my focus will not be refuting views like those held by Fodor and Chomsky, but on evaluating McDowell’s Wittgensteinian response to the rule-following problem.)

⁴For details on Wright’s claim that the manifestation problem merely presents a challenge to the realist, rather than an argument demonstrating the falsity of realism see (Wright, 1989/1993b), particularly p. 246.

⁵Wittgenstein, of course, does not organise his claims into clear step-by-step arguments, so in this summary I shall be relying on a combination of McGinn’s exposition of Wittgenstein in *Wittgenstein on Meaning* (McGinn, 1984) and Kripke’s in *Wittgenstein on Rules and Private Language* (Kripke, 1982).

sentence ‘That is a square’ is true we would look at the object in the world, and compare it to our square mental image to check if they matched. Wittgenstein accepts that this story may occasionally be phenomenologically accurate, but argues that there are three problems with it:

The “grammatical” problem: Wittgenstein’s first argument against this view of meaning is based on the fact that experience is temporal; for example, we are perfectly happy to talk of a toothache starting at 3pm and ending at 4pm, and of it being continuous or sporadic over a period of time. Understanding is not so easily clockable in these ways. (It’s true that the manifestations of our grasp of a meaning do happen at particular times, but this is not significant here because *possession* of meanings is the key concept; it would seem strange to say someone only means, say, plus by ‘+’ when they are manifesting their understanding by using the word.) In addition, Wittgenstein points out that it is appropriate to talk about the intensity or quality of an experience — we distinguish between pains being throbbing, stabbing or dull, etc. —, but there is no parallel of this when talking about understanding. According to McGinn (1984, p. 5–6), Wittgenstein uses cases like these to argue that there is a “grammatical” difference between understanding and experiences, because it would be “ungrammatical” to talk about understanding in the same kind of ways in which we talk about experiences. He concludes that the difference is that experience is an occurrent concept, whereas understanding is more dispositional (i.e. the concept of experience is the concept of something that happens at a particular time and the concept of understanding is not).

Mental images aren’t necessary: Wittgenstein’s other two arguments against the experiential view of meaning are based on introspection of the phenomena of meaning something. The first is that they cannot be *necessary* for understanding, because of the simple fact that in the vast majority of cases we successfully mean things without any image coming before our minds — we can use words correctly while picturing something completely different, and we can even use words without picturing anything at all.

Mental images aren’t sufficient: He argues that mental images cannot be *sufficient* for meaning because their correct interpretation is essentially indeterminate:

I see a picture; it represents an old man walking up a steep path leaning on a stick. — How? Might it not have looked just the same if he had been sliding downhill in that position? Perhaps a Martian would describe the picture so. (Wittgenstein, 1953/2001, p. 54, fn. a).

As Kripke (1982, p. 42) notes, one response to this indeterminacy would be to claim that what comes to mind is not only the picture, but a method of interpreting it,

which shows us how to correctly interpret the picture. But how are we to explain our understanding of this interpretation? Surely we can only understand how to use the interpretation because of our understanding of its meaning, and so our use of it must rely on a further picture, which also would need to be interpreted correctly. It seems that for any mental image to give us a determinate meaning it needs to be interpreted, but we also need to know the meaning of the interpretation. If mental images are how we know meaning then to know the meaning of the interpretation we'd need a further mental image, which in turn would require interpretation, generating an infinite regress. (I'll go into this argument in much greater detail in §2.1.2 of chapter 2.)

Wittgenstein also attacks dispositional accounts of meaning, which are those that hold that what we mean by a sentence is determined how we are disposed to behave when presented with it. This is intended to respond to Wittgenstein's rule-following argument,⁶ and explain what determines the meanings of our words in cases where we are yet to use them. For example, according to the dispositional account, even though we may have never worked through the addition problem '68 plus 57' there is still a fact about what we mean by 'plus' in that situation because we would be disposed to give the answer 125. Kripke (1982, pp. 22–32) goes over several problems with this view, including the following four:

Normativity: The core problem with the dispositional view is that it doesn't account for the normativity of meaning. Kripke argues that, in contrast to dispositions, "The relation of meaning and intention to further action is *normative*, not *descriptive*." (Kripke, 1982, p. 37). The meaning of '+' should provide *justification* for responding '125' to '68+57', and should determine that this answer is correct and any others are wrong. Dispositions merely *describe* how I will respond, they don't tell me how I *should* respond. This is problematic because meaning is normative — there are right and wrong ways to use language, and dispositions struggle to provide this normative force. Instead, dispositions merely describe what seems to me to be right, and as Wittgenstein argues, if "whatever is going to seem right to me is right" then "that only means that here we can't talk about right" (Wittgenstein, 1953/2001, §258), and so we lose grip on the normativity of meaning.

The finitude of dispositions: The problem here is that our dispositions are finite; we would die before we were able to work out the solution to really large sums. But it would seem odd indeed if 'plus' didn't mean anything for sums greater than those we'd be able to compute in a lifetime; surely the meaning of 'plus' should determine

⁶The rule-following argument will be the focus of next two chapters; in this section I'll just give the briefest of sketches to demonstrate Wittgenstein's arguments for his view of language.

what the correct answers are for all calculations, regardless of the size of the numbers they contain.

Mistakes: We may have the disposition to make mistakes, we may happen to be disposed to get sums wrong sometimes. For example, I may have the disposition to respond ‘5’ when asked to compute ‘ $68+57$ ’, but we wouldn’t say that I *mean* something different by ‘+’ than my peers who would get the sum right, just that I make a particular systematic mistake. We wouldn’t say that according to *my* meaning of ‘+’ that ‘5’ is the correct answer, instead we’d say that I got the sum wrong. Or alternatively, consider coming across a horsey-looking cow on a dark night.⁷ I may well be disposed to call such an animal a ‘horse’, despite it actually being a cow. We would want to say by calling the cow a ‘horse’ I’d made a mistake; ‘horse’ should only be used to refer to horses, not to cows in bad lighting conditions that just happen to look a little horsey. But if the meaning is just my dispositions then my word ‘horse’ would mean *horse or cow on dark nights*, which is clearly the wrong result. The problem here is that dispositional accounts have difficulty in distinguishing between performance and competence; it’s hard for them to draw the distinction between answers which merely seem right and those which actually are right. This is effectively just another way of putting the normativity point; meanings should tell us what we *ought* to do in any particular case, not merely describe what we would happen to do for each case.

Knowledge of our meaning: In addition to these three, Wright (1998/2000) argues that another difference between meaning and dispositions, which is central to Wright’s reading of Wittgenstein, is that they have different epistemic characteristics. Third person and first person perspectives are epistemically equal for dispositions, since our knowledge of our dispositions is based on inferences from our behaviour, which a well placed external observer could have equally good knowledge of. In contrast, for knowledge of meanings first person and third person perspectives are significantly different; a speaker can have superior knowledge of what they mean by a word than someone else would, since their knowledge of what they themselves mean seems to be non-inferential, whereas from the third person perspective our knowledge of what other people mean is based on inferences. However, this reading of Wittgenstein is controversial. For example, McDowell (2000) argues that our knowledge of other people’s meaning isn’t inferential and that when things go right we can simply hear what someone means by a word when they say it.

⁷We’ll see this example again in §2.3 of the next chapter when I evaluate the dispositional account of meaning that Wright develops in *Wittgenstein on the Foundations of Mathematics* (Wright, 1980). Wright quickly went on to drop this account and move on to a more sophisticated approach, which we’ll also see in chapter 2.

According to McGinn, Wittgenstein concludes that we should abandon the attempt to reduce meaning to dispositions and the view of meaning as a hidden inner state which our use of signs merely manifests, and instead hold that understanding *consists* in our use of signs. This leads him to the view that we should instead think of understanding as more like an ability:

We get the grammar of ‘understands’ right if we classify it along with being able to play chess, sing a tune, appreciate a piece of music, employ a tool — instead of assimilating it to having a visual experience, imagining a face, having a feeling of confidence, thinking up a hypothesis. (McGinn, 1984, p. 31).

The other aspect of the Wittgensteinian view of meaning which is central to the manifestation argument is the claim that our meaning should be associated only with the abilities which we are capable of manifesting publicly. Dummett justifies this claim as follows:

To suppose that there is an ingredient of meaning which transcends the use that is made of that which carries the meaning is to suppose that someone might have learned all that is directly taught when the language of a mathematical theory is taught to him, and might then behave in every way like someone who understood that language, and yet not actually understand it, or understand it only incorrectly. But to suppose this is to make meaning ineffable, that is, in principle incommunicable. If this is possible, then no one individual ever has a guarantee that he is understood by any other individual; for all he knows, or can ever know, everyone else may attach to his words or to the symbols which he employs a meaning quite different from that which he attaches to them. (Dummett, 1973/1978a, pp. 217–218).

Dummett’s argument here revolves around the point that if meanings went beyond what was manifestable publicly then meaning would involve something which goes beyond what we’re able to communicate, since, as he says earlier in the same paper, “An individual cannot communicate what he cannot be observed to communicate” (Dummett, 1973/1978a, p. 216). And clearly what we are able to mean cannot go beyond what we can communicate.

1.1.2 The truth-conditional conception of understanding

According to Frege, when giving an account of the meaning of a sentence we must distinguish between sense and force. This means that our theory of meaning will include the following two parts:

a core theory, which determines the conditions for the application to sentences of the language, of some meaning-relevant property (candidates being, for

instance, truth or warranted assertability); and a supplementary part, the theory of force, which gives an account, in terms of the central notion of the core theory, of what it is to effect each of the various possible types of speech act. (McDowell, 1981/1998b, p. 318).

Consider the following three sentences:

- The window is open.
- Open the window!
- Is the window open?

Each of these sentences is conventionally used with a different force; the first is conventionally used to make an assertion, the second an order, and the third a question. The literal meaning of the second and third is in part a function of the thought expressed by the first.

According to Frege, the thought expressed by a use of a sentence is a function of the sense or literal meaning of that sentence and the force with which it is used. The literal meaning of all three is given in terms of the truth-conditions of the first, and an indication of the force with which they are conventionally used.

According to Frege's view of meaning, the thought expressed by a sentence is a function of its sense and force. The sense of all three is given by the truth-conditions of the first, which is then combined with their force to give the meaning. So, for example, the meaning of the third is a result of the combination of the truth-conditions of 'The window is open' with the interrogative force. According to this view, truth-conditions are at the very core of our theory of meaning, since they constitute the sense of all sentences, which is an essential part of the thought expressed. (In contrast, some anti-realist views, such as those endorsed by Dummett in his earlier work⁸ claim that we should replace truth with warranted assertability at the core of our theory of meaning, and we should focus on an account of the situations in which we would be justified in asserting any sentence.) McDowell argues for truth playing this central role as follows:

The basis of the truth-conditional conception of meaning, as I see it, is the following thought: to specify what would be asserted, in the assertoric utterance of a sentence apt for such use, is to specify a condition under which the sentence (as thus uttered) would be true. (McDowell, 1987/1998e, p. 88).

Clearly, at least according to the above argument, to understand the content of an assertion it is essential that we understand what it is for the assertion to be true or false, and so our theory of meaning must be able to give an account of what it is for sentences that are used to make assertions to be true or false.

⁸Such as "What is a Theory of Meaning? (II)" (Dummett, 1976).

1.1.3 Evidence-transcendent truth

This is the claim that it is possible for sentences to have potentially evidence-transcendent truth-conditions; they can be true or false even if we have no evidence for their truth-value. Consider the following two sentences:

- James II suffered a migraine on the afternoon of his 32nd birthday.
- Goldbach's conjecture: Every even number greater than two is the sum of two primes.

These are examples of sentences that may have evidence-transcendent truth-conditions because we do not currently know their truth-value, and nor do we know an effective procedure for discovering their truth-value.

For example, we do not currently have access to any historical evidence, nor do we know of a finite procedure that is guaranteed to give us such evidence, which would justify us believing that James II did or did not suffer a migraine on that particular birthday, and for all we know no such evidence exists; this shows that the sentence's truth-conditions could be evidence-transcendent. Similarly, we do not know whether every even number greater than two is the sum of two primes; there is currently no mathematical proof to show that it is true or a counterexample to show that it is false, and we don't know of any finite process which is guaranteed to generate such a proof or counterexample.

A core concept here is decidability. A decidable sentence is one whose truth-value we currently know, or for which we have an effective procedure with which to decide its truth-value, where an effective procedure is one that will:

in finite time, put [us] into a position in which [we] can recognize whether or not the condition for the truth of the sentence is satisfied. (Dummett, 1976, p. 81).

Undecidable sentences are those for which we lack such an effective procedure, and so we lack a finite process that we know will provide us with a justified belief about the sentence's truth-value. The two above sentences are significant precisely because they are undecidable.

Despite the fact that these sentences are undecidable, and that we don't know if we'll ever be able to discover whether they are true or false, most of us still have the realist's intuition that they must be one or the other. It is this intuition — that a sentence's truth can go beyond the availability of evidence for it — which this third realist commitment identifies.

It's important to note that a realist about, say, the past, doesn't necessarily also have to be a realist about other domains, such as mathematics. It's perfectly coherent to think that there are undecidable sentences about the past (such as the first example

above) with determinate truth values, and yet also think that there's no reason to believe that undecidable mathematical sentences have determinate truth-values. However, such a nuanced position is irrelevant to the manifestation argument, which aims to support anti-realism globally, across all domains.

It's also important to emphasise that Dummett's notion of undecidability is epistemic. In areas such as number-theory a sentence is decidable if it is derivable within a particular formal system. This is a non-epistemic notion, and it only makes sense to talk about decidability with respect to a particular system, and not decidability simpliciter. But in the paper "Undecidability in Anti-Realism" (Shieh, 1998, pp. 325–326) Shieh points out that when Dummett talks about decidability he doesn't do so in reference to any particular axiom system. Instead, Shieh argues that for Dummett a statement's decidability depends upon whether *we know* an effective procedure for recognising its truth-value. For example, for all we know it may well be possible to generate a proof of Goldbach's conjecture using our current mathematical apparatus, or we may just happen to stumble upon a counterexample by chance. But we do not know (or have reason to believe) that such a proof is possible, and so while we continue to lack reason to believe that such a proof exists Goldbach's conjecture will continue to count as undecidable.⁹

Bivalence and realism

Dummett uses the notion of bivalence to clarify the role the potential for evidence-transcendent truth plays in the debate between realists and anti-realists. Bivalence is the view that all well-formed sentences in a particular discourse are determinately either true or false. If we combine bivalence with the claim that some of the sentences in the discourse are undecidable then realism must be true of that discourse. This is because there will be sentences that have a determinate truth-value, but we know of no way to determine what their truth-value actually is; truth (in the discourse) potentially outruns our ability to discover it. Anti-realism consists in the rejection of this claim, and the assertion that we have no justification for supporting either bivalence across the board¹⁰ or for thinking that the facts may go beyond what we can know.

It is essential to point out that this doesn't mean that anti-realists believe that undecidable sentences have some third truth-value, neither truth nor falsity. Dummett clarifies this point as follows:

⁹For more on Dummett's use of the notion of undecidability it is worth looking at not only Shieh's entire paper, but also the appendix on decidability in Weiss' book *Michael Dummett* (Weiss, 2002, pp. 177–178).

¹⁰The anti-realist doesn't claim that there are no areas of discourse in which bivalence is true, for example they would be happy to accept bivalence for elementary — quantifier-free — arithmetic. Instead they are rejecting the realist's support for bivalence in areas of discourse which include undecidable sentences.

Of any statement we can never rule it out that we might subsequently come upon something which justified asserting or denying it, and therefore we are not entitled to say of any specific such statement that it is neither true or false: but we are not entitled either to say in advance that it has to be either one or the other, since this would be to invoke notions of truth and falsity independent of our recognition of truth or falsity, and hence incapable of having been derived from the training we received in the use of these statements. (Dummett, 1969, p. 246).

As we can see, the anti-realist's claim is merely that we are not justified in endorsing bivalence — we are not justified in asserting that all well-formed sentences, even the undecidable ones, must be either true or false. And their rejection of bivalence involves nothing more than that.

Another important clarificatory note is that when refusing to endorse bivalence, the anti-realist is intending to do so independently of problems surrounding vagueness. As Wright (1993a, p. 4) notes, vagueness is a pervasive feature of most non-mathematical talk, and so we would expect even the realist to reject global bivalence for most discourses. The realist position is interesting because it involves holding that sentences such as those about James II's headaches have a determinate truth-value, despite our lack of evidence; the position is made no less interesting if they don't think that sentences such as 'Those grains of sand form a heap.' can sometimes fail to be determinately true or false. Instead, the anti-realist's rejection of bivalence is meant to apply to sentences that are non-vague, and so, according to Dummett, the realist position is one which holds that all non-vague sentences have a determinate truth-value.

McDowell (1989/1998g, p. 355) argues that Dummett's view of the role bivalence plays in the realism / anti-realism debate has problems beyond those surrounding vagueness. He acknowledges that bivalence implies realism, because if you believe that necessarily all well-formed sentences are true or false (bivalence), then reality must be determinate beyond our ability to recognise it (realism). But he points out that the converse isn't true, and that it's perfectly possible to reject bivalence and still be a realist. This is because technically if all that's at the core of the realist position is the claim that reality sometimes goes beyond what we are able to verify, then all the realist is committed to is the claim that there are *some* well-formed sentences with potentially verification transcendent truth-conditions which are determinately true or false. There's no need for the realist to endorse the much stronger claim of bivalence and think that necessarily *all* well-formed, non-vague, sentences are true or false. When making this point, McDowell explicitly notes that this is not meant to be connected to problems surrounded vagueness, but he's hesitant to put forward a positive story about a realist view which would actually endorse this kind of position; his goal here is merely to point out that there is at least scope for a position of this kind and that therefore bivalence shouldn't be seen as the

central focus of the debate.¹¹ Fortunately, McDowell’s criticisms of Dummett on this issue have no significant impact on the manifestation argument, since the core of the issue is about whether there are *any* well-formed non-vague sentences which lack a determinate truth-value, and so for the rest of this chapter I can simply ignore this detail.

1.2 The significance of semantic realism

Before getting into the meat of Dummett’s arguments against realism, it would be worthwhile to briefly consider a more fundamental criticism of his entire approach. Central to Dummett’s philosophical work is the claim that metaphysical debates about realism are best viewed as semantic debates, and therefore best solved by turning to philosophy of language. When talking about the realism debate in the context of mathematics, Dummett claims:

we cannot . . . *first* decide the ontological status of mathematical objects, and then, with that as a premise, deduce the character of mathematical truth . . . Rather, we have first to decide on the correct model of meaning . . . and then one or the other picture of the metaphysical character of mathematical reality will force itself upon us (Dummett, 1973/1978a, p. 229).

And in the paper “Truth” Dummett states that: “realism is a semantic thesis, a thesis about what, in general, renders a statement in the given class true when it is true” (Dummett, 1963/1978b, p. 145). However, this view has met serious opposition. A central goal of Devitt’s book *Realism and Truth* (Devitt, 1991b) is to undermine this view of the primacy of philosophy of language:

It is a mistake to start building a metaphysics from epistemology or semantics. The realism issue should be settled first. Failing to do so is one of the most pervasive and serious aberrations of the realism debate . . . To suppose that we can derive the right metaphysics from epistemology or semantics is to put the cart before the horse. (Devitt, 1991a, p. 56–57).

Clearly, the debate about the relationship between metaphysics and philosophy of language is a substantial debate in its own right, and a thorough response to the critics of Dummett’s entire approach to philosophy is well beyond the scope of this chapter. Instead I aim merely to sketch one reason for taking Dummett’s arguments against se-

¹¹In a footnote on this topic McDowell does suggest that the realist about ascriptions of sensations to others may not wish to endorse bivalence about that domain, but he doesn’t really flesh the example out with any detail (McDowell, 1981/1998b, p. 355, fn. 26). When writing on this Wright accepts that McDowell is correct to point out that realism doesn’t necessarily imply bivalence, although he too is unsure about when a realist may actually want to hold such a position (at least for reasons other than those involving vagueness) (Wright, 1980/1993c, p. 85).

mantic realism seriously, despite flaws in his own arguments for the primacy of philosophy of language.¹²

Dummett’s argument for the primacy of semantics rests upon the claim that debates between realists and their opponents have been guilty of depending upon metaphors which are hopelessly ambiguous:

We are, after all, being asked to choose between two metaphors, two pictures. The platonist metaphor assimilates mathematical enquiry to the investigations of the astronomer: mathematical structures, like galaxies, exist, independently of us, in a realm of reality which we do not inhabit but which those of us who have the skill are capable of observing and reporting on. The constructivist metaphor assimilates mathematical activity to that of the artificer fashioning objects in accordance with the creative power of his imagination. Neither metaphor seems, at first sight, especially apt, nor one more apt than the other: the activities of the mathematician seem strikingly unlike those either of the astronomer or the artist. What basis can exist for deciding which metaphor is to be preferred? How are we to know in which respects the metaphors are to be taken seriously, how the pictures are to be used? (Dummett, 1973/1978a, p. 229).

However, other than his vivid prose, Dummett has little in the way of argument for the claim that metaphysical debates will necessarily be clouded by metaphor. Unarguably metaphors have often clouded metaphysical debates, but that fact alone doesn’t show that they must. Consider the following formulation of realism, offered by Miller:

Common Sense Realism: Tokens of most current observable common-sense and scientific physical types objectively exist independently of the mental; they possess some properties which may pass altogether unnoticed by human consciousness (Miller, 2003b, p. 199).

Miller argues that there is nothing metaphorical about this, or at least if there are ineliminable metaphors hidden away within formulation such as this, that is something it would take substantial argument to show, and Dummett has failed to provide any such argument.

However, Miller continues by arguing that this doesn’t mean that when considering debates about realism itself we can simply ignore Dummett’s arguments against semantic realism. He points out that a successful realist position needs to do more than argue for a realist metaphysics; it must also provide a metaphysics which fits into an entire “realistic worldview” which involves compatible, and coherent, epistemic and semantic elements:

What is a worldview? A worldview consists of at least a *metaphysics* (an account of what there is and its nature in general), an *epistemology* (an account

¹²Much of the work in this section draws heavily from Miller’s paper “The Significance of Semantic Realism” (Miller, 2003b), which is a much more thorough exploration of these issues. Miller also explores these issues in “Realism, and Anti-realism” (Miller, 2006).

of how we can possess knowledge of the objects and properties included in the metaphysics), and a *semantics* (an account of how we can talk and think about the objects and properties included in the metaphysics). (Miller, 2003b, p. 206, emphasis in original).

What's crucial about Miller's claim here is that, no matter how good our realist metaphysical picture may appear when considered in isolation, if we can't make it fit with a compatible semantics then it fails as a picture of a world which we are able to talk and think about. (And, if we are unable to account for our ability to talk and think about the objects and properties proposed by or metaphysics, then it's hard to see how we can even be said to have been able to argue for that view of reality in the first place.) Once we consider the metaphysical issues in this light, the significance of Dummett's work becomes apparent.

Dummett's arguments against semantic realism, such as the acquisition and manifestation arguments, aim to show that our understanding of sentences doesn't consist in a grasp of potentially evidence-transcendent truth-conditions. Anti-realists typically do this in one of two ways: by completely rejecting the idea that understanding declarative sentences consists in a grasp of their truth-conditions (rejecting **2**, above); or, by arguing that truth is an epistemically constrained notion (the rejection of **3**). These two conclusions have significantly different consequences for our theory of meaning, but as far as evidence-transcendence is concerned they are on a par in that neither leaves room for the idea that we can understand sentences to have determinate truth-values beyond those which we could have evidence for.

Miller points out that a semantic theory which claims that our sentences can't have evidence-transcendent truth-conditions can't be incorporated into a worldview that also includes common sense realism as its metaphysical component. This is because common sense realism requires us to be able to make sense of the idea that things may "possess some properties which may pass altogether unnoticed by human consciousness" (Miller, 2003b, p. 199). But if the sentences we use when thinking and talking about reality cannot have truth-conditions beyond what we can potentially have evidence for, then how is it possible for us to make sense of the realist's position? This means that, if Dummett's arguments against semantic realism work, then they have a profound influence on the type of plausible worldview which we can construct. And, unless the realist can provide a successful refutation of Dummett's argument, then they must either: produce an alternative realist conception of linguistic understanding than the truth-conditional one, which is invulnerable to Dummett's attacks; or, drop their support for realism.

Of course, there may well be alternative realist conceptions of linguistic understanding which are compatible with the possibility of evidence-transcendent facts. For example, Devitt (in (Devitt, 1991b) and (Devitt, 1991a)) seeks to undermine Dummett's argument by rejecting the idea that understanding a sentence consists in knowledge of its truth-

conditions (well, in fact he rejects that idea understanding a sentence involves knowledge at all). However, evaluating such a move is beyond the scope of this chapter. Instead, I shall focus on attempts to meet Dummett’s argument head on, and show that his arguments fail to demonstrate the incompatibility of the three realist commitments, and that our sentences can have evidence-transcendent truth-conditions.

1.3 The acquisition argument

The manifestation argument was initially inspired by the discovery of shortcomings in a parallel argument against realism developed by Dummett, the acquisition argument.¹³ This argument focuses around the fact that our linguistic abilities are necessarily learnt by training from other speakers, and so cannot possibly go beyond what we can be taught to understand. Dummett argues that semantic realism is committed to a view of our understanding of sentences that goes beyond what we could be taught, and so must be false. This argument takes the form of a *reductio*, with the initial premises being along these lines:

- A. Our understanding of sentences is acquired by training.
- B. Understanding a declarative sentence consists in grasp of its potentially evidence-transcendent truth-conditions; grasp of the truth-conditions of a sentence may require an understanding of how it could be undetectably true or false.

The conflict between these two premises arises because the training upon which our understanding of language relies is restricted to cases where we are capable of recognising a state of affairs that warrants asserting that the truth-condition obtains. For example, one would be taught to understand the sentence ‘This is square’ in the presence of a square object whose squareness the language learner would be capable of recognising. But if this is so, then how is it we are able to acquire a grasp of meanings which go beyond our recognitional abilities in the ways envisaged in **B**? Consider the two sentences used above to exemplify evidence-transcendent truth:

- James II suffered a migraine on the afternoon of his 32nd birthday.
- Goldbach’s conjecture: Every even number greater than two is the sum of two primes.

¹³This summary of the acquisition argument owes much to Miller’s paper “What is the Acquisition Argument?” (Miller, 2003c). For examples of Dummett and Wright’s own exposition of the argument see papers such as Dummett’s “Truth” Dummett (1959), and Wright’s Introduction to *Realism, Meaning and Truth* Wright (1993a). Hale’s “Realism and its Oppositions” (Hale, 1997) is also useful on this topic.

These sentences both have potentially evidence-transcendent truth-conditions, and so we do not know if we'll ever have the ability to recognise their truth or falsity. But the anti-realist argues that the conditions to which we are exposed when we learn the meaning of a sentence are conditions which we are guaranteed to be able to recognise to obtain when they obtain — so how do we acquire the concept of truth-conditions that can obtain in the absence of such a guarantee?

Wright puts the argument as follows:

How are we supposed to be able to *form* any understanding of what it is for a particular statement to be true if the kind of state of affairs which it would take to make it true is conceived, *ex hypothesi*, as something beyond our experience, something which we cannot confirm and which is insulated from any distinctive impact on our consciousness? (Wright, 1993a, p. 13, emphasis in original).

The anti-realist argues that in response to the conflict between **A** and **B** we must reject **B**, and instead endorse an evidentially constrained view of sentence understanding. The logical form of the argument would look something like this:

- **A**
- If **A** then \neg **B**
- \therefore \neg **B**

1.3.1 The realist's response

The typical realist response to the argument is to point out that our understanding of language is compositional; we understand the meaning of sentences by grasp of their constituents and mode of syntactic combination (i.e. by the way those constituents are put together). For the vast majority of the sentences that we are able to understand we didn't hear them when learning to speak, and yet we are able to understand them because we understand their constituents, and how to understand the meaning when their constituents are put together. The realist claims that sentences with potentially evidence-transcendent truth-conditions (such as the two above) are no different; just because we were never explicitly taught to understand them in a situation where we would be warranted in asserting that their truth-condition obtains, that is no reason to claim that we don't actually understand their truth-conditions. For example, we understand other sentences involving James II from history books which warrant assertions about him such as 'James II is dead', we understand sentences involving migraines such as 'Mary has a migraine', and we understand sentences about what it is for something to happen to someone on their 32nd birthday such as 'Bill went to the bar on his 32nd birthday'. So why wouldn't we understand the truth-conditions of 'James II suffered a migraine on

the afternoon of his 32nd birthday.’? Why would the fact that the sentence has potentially evidence-transcendent truth-conditions block our ability to understand its meaning though grasp of its constituents?

A possible, but flawed anti-realist response to this argument would be to claim that this begs the central question. Of course, the anti-realist continues, our understanding of language is compositional, but what justifies the realist’s claim that we understand what it is for the controversial sentences to have *evidence-transcendent* truth-conditions? Perhaps we have an evidence-constrained understanding of truth, and that our understanding of what it is for the related sentences — ‘James II is dead’ etc. — to be true or false is restricted to cases where are warranted in making an assertion one way or the other. Clearly we’ve only been trained in cases where we have sufficient evidence to warrant an assertion, so there’s no evidence there that we have an evidence-transcendent understanding of truth. So, this anti-realist argument concludes, the compositionality of meaning provides us with no evidence that the realist’s view of sentence understanding **B** is correct, instead of an evidence-constrained alternative.

The problem with this response to the realist is that the acquisition argument attempted to prove the falsity of realism **B** by arguing that if **A** then \neg **B**. All the realist has to do to successfully respond to this argument is to show that **A** is consistent with **B** and that is what the compositionality considerations demonstrate.

1.4 The manifestation argument

The acquisition argument attempted to demonstrate a conflict between realist accounts of sentence understanding and the fact that our understanding can’t go beyond what can be acquired by training. The manifestation argument parallels this, but the difference is that the problem for realism instead comes from the claim that there is *nothing more* to our understanding than what is publicly manifestable. Remember the Wittgensteinian slogan: “Meaning cannot transcend use” (McDowell, 1981/1998b, p. 344). If realism is true then that must be manifest in our use of language, otherwise we’d be guilty of thinking that there’s more to our understanding than what is manifest in the use we make of sentences. Here’s Dummett:

It is, in fact, plain that the knowledge which is being ascribed to one who is said to understand the sentence is knowledge which transcends the capacity to manifest that knowledge by the way in which the sentence is used. The [realistic] theory of meaning cannot be a theory in which meaning is fully determined by use. (Dummett, 1973/1978a, p. 224–225).

It is because the manifestation argument relies on the claim that there is nothing more to our understanding of language than what is publicly manifestable that the realist use

of compositionality in the response to the acquisition argument won't work here.¹⁴ As we saw in the previous section, reflecting on the compositionality of meaning reveals that the realist understanding of truth is compatible with how we acquire linguistic understanding. More is needed to respond to the acquisition argument because it requires actual manifestation of the claim that the notion of truth involved is evidence-transcendent. Our use of the constituent parts of the controversial sentences about Goldbach's conjecture and James' headache in more straightforward cases does nothing to manifest that our understanding is evidence-transcendent. So the compositionality of meaning can't be used (at least straightforwardly) to respond to the manifestation argument.

The most mature form of the manifestation argument was developed by Wright, and is best understood as an argument for the incompatibility of three realist commitments:

1. **The Wittgenstein conception of linguistic understanding:** Linguistic understanding consists in a collection of practical abilities, and there is nothing more to our understanding than what can be manifested publicly; "Meaning cannot transcend use" (McDowell, 1981/1998b, p. 344).

For the purposes of *reductio* assume the other two realist commitments:

2. **The truth-conditional conception of understanding:** Understanding a declarative sentence consists in grasp of its truth-conditions.
3. **Evidence-transcendent truth:** Truth is unconstrained by the availability of evidence; grasp of the truth-conditions of a sentence may require an understanding of how it could be undetectably true or false.

1.4.1 Core and neighbourhood abilities

There are many different interpretations of the manifestation argument, and many commentators distinguish between two different strengths of conclusion which Wright has argued for.¹⁵ For the purposes of clarity I shall start out by explaining a slightly simplified version of Wright's argument, which will hopefully make explaining the need for the more nuanced aspects more straightforward.

¹⁴Or at least, a response to the problem involving compositionality won't be so straightforward. For an in-depth analysis of whether compositionality can be used to respond to the manifestation argument see Byrne's paper "Compositionality and the Manifestation Challenge" (Byrne, 2005).

¹⁵For example, Miller (2002) draws a distinction between a "strong" and a "weak" version of the argument, and Byrne (2005) draws a parallel difference between the manifestation "argument" and the manifestation "challenge". I don't believe there is anything more significant than a verbal difference between the ways Byrne and Miller draw this distinction, but the verbal difference is important to keep in mind, since Miller uses the word "challenge" when talking about both versions of the argument, whereas Byrne uses it only to refer to the weaker. In this chapter I will follow Byrne's way of talking about the distinction.

How should we understand premise **1**? What abilities must we be able to manifest in order to count as understanding a sentence? Well, if, as **2** asserts, understanding declarative sentences consists in grasp of their truth-conditions, then understanding such sentences must require a grasp of what it would be for them to be true or false. This means that the practical abilities that must be manifestable by speakers who understand declarative sentences must consist in grasp of their truth-conditions.

In order to focus more clearly on this ability Wright draws the distinction between the ability to recognise a sentence's truth-value and mere "neighbourhood abilities".¹⁶ Consider the understanding of sentences involving taste, such as 'This is sweet', or 'This is bitter' (in a case where we are talking about items whose taste we have easy access to). The neighbourhood abilities involved in understanding such sentences will include being able to appraise evidence for them, and being able to recognise inferences to and from such sentences, being able to use them in justifying ascriptions of propositional attitudes to people (such as why Jones puts that substance rather than another one in his tea).¹⁷ Although such abilities are important, Wright (1993a, pp. 17–18) argues that neighbourhood abilities should not be the focus of the current debate; instead we should focus on the "core abilities", which are abilities like being able to "recognize the taste of the samples by placing them in one's mouth, and thereby verify or falsify descriptions of their taste" (Wright, 1993a, p. 17). According to Wright, the core abilities — the abilities involved in recognising a sentence's truth-value — can be *identified* with knowledge of the sentence's truth-conditions, as in the case of declarative sentences involving taste, because:

recognizing the taste of something by tasting *is* recognizing that in virtue of which the true description of its taste is true (Wright, 1993a, p. 17, emphasis in original).

In summary, according to **1** understanding consists in a publicly manifestable practical ability. What is this ability? Well, **2** asserts that understanding declarative sentences consists in grasp of their truth-conditions. So, at least in the case of declarative sentences, this means that the ability that constitutes understanding must involve grasp of the sentence's truth-conditions. And, Wright argues, what better candidate could there be for the ability that constitutes grasp of a sentence's truth-conditions than the very ability to recognise the truth or falsity of the sentence? So he concludes that this ability has a "plausible claim to be regarded as central, or even constitutive of understanding" (Wright, 1993a, p. 18). This gives us the next step in the argument:

¹⁶"Neighbourhood abilities" is a phrase originally coined by Blackburn, which has been adopted by Wright.

¹⁷These examples are all taken from Wright (Wright, 1993a, p. 17).

4. Understanding a declarative sentence's truth-conditions requires the manifestable ability to recognise whether its truth-condition obtains.

It's important to note at this stage that **4** does not necessarily follow from **1** and **2**, it is merely a candidate for the ability required by **1** which is compatible with **2**. So far we've seen no argument to show that there couldn't be other abilities which play this role, but Wright has at least shown it to be a highly plausible candidate.¹⁸

1.4.2 The argument

Consider the examples of potentially evidence-transcendent sentences used above:

- James II suffered a migraine on the afternoon of his 32nd birthday.
- Goldbach's conjecture: every even number greater than two is the sum of two primes.

These are classic examples of the sort of sentences with evidence-transcendent truth-conditions to which **3** refers. And according to **3** for sentences such as these two the following will be true:

5. Grasp of the truth-conditions of certain sentences may require understanding of how they can be undetectably true or false.

Clearly if a sentence is undetectably true or false, then we may¹⁹ be unable to recognise its truth-value, so from **5** we get:

6. We can grasp the truth-conditions of declarative sentences whose truth-value we may be unable to recognise.

But from **4** it follows that:

7. We do not grasp the truth-conditions of a declarative sentence if we lack the ability to recognise its truth-value.

This gives us the desired conclusion:

8. Reductio, **6** and **7** contradict. so we must reject one of the initial premises.

¹⁸In singling out **4** as a step in a the manifestation argument I am following Byrne's exegesis of the argument in "Compositionality and the Manifestation Challenge" (Byrne, 2005). In that paper he refers to **4** as "the natural supposition", which he defines as the claim that "A speaker's understanding of a declarative sentence incorporates the ability to recognize (in appropriate circumstances) whether its truth condition obtains." (Byrne, 2005, p. 109). I am also following Byrne in my claim that Wright puts **4** forward as a plausible candidate for the ability required by **1** and **2**, and doesn't provide a general argument for it.

¹⁹It's only possible to assert that we *may* be unable to recognise the truth-value of undecidable sentences, and not that we *can't* because, as we saw above, undecidability is an epistemic notion. A sentence is undecidable if we don't currently know its truth-value and don't know of a finite procedure which is guaranteed to provide warrant for asserting the sentence to be true or false. That doesn't rule out that we may discover there to be such a procedure, or accidentally stumble upon evidence one way or the other.

1.4.3 The manifestation challenge

However (as alluded to after defining **4** above), there's a serious problem for this argument, in that **4** doesn't *necessarily* follow from **1** and **2**. Remember, **4** was defined as:

4. Understanding a declarative sentence's truth-conditions requires the manifestable ability to recognise whether its truth-condition obtains.

Byrne (2005, pp. 110–112) points out that Wright's argument for the association between understanding and recognitional abilities focuses around the case of sentences such as 'this is sweet', or 'this is bitter'. Wright here is clearly not aiming to make a specific point about our understanding of taste, and since these sentences are fairly paradigmatic cases of decidable sentences it is probably safe to generalise Wright's conclusion here to all decidable sentences. But **4**, at least as defined in the above formulation of the argument, applies to all declarative sentences, decidable or undecidable, and it's not at all clear that Wright's argument using taste examples gives us any reason to think that the ability required to understand *undecidable* sentences must also be the ability to recognise the obtaining of their truth-condition.²⁰ Indeed, such a claim seems unavailable, since we lack that very ability in the case of undecidable sentences. This is problematic for the anti-realist's argument, since if we restrict **4** to only apply to decidable sentences then it will be useless for providing the required reductio.

There still remains a serious problem for the realist though, as Wright points out:

A realist who accepts the Wittgensteinian premise seems to have no general candidate for an ability, common to all cases of statement-understanding, of which 'knowledge of truth-conditions' is a well-motivated description. (Wright, 1993a, p. 18).

The problem here is that if we restrict **4** to apply only to decidable sentences then we are left with no account of the abilities that constitute our understanding of undecidable sentences and constitute grasp of their truth-conditions. According to **1** (the Wittgensteinian conception of linguistic understanding), there is nothing more to our understanding than the practical abilities that we are able to manifest, so the realist still needs to give us some account of what abilities we possess which enable us to manifest that our understanding is realist. Otherwise we must conclude that we lack such an understanding.

So, the realist needs an alternative to **4**; they need a description of the manifestable abilities that constitute our understanding of sentences with potentially evidence-transcendent truth-conditions. And if the abilities that constitute understanding are not the abilities to recognise the obtaining of the sentence's truth-conditions (i.e. they are not those specified by **4**), then what are they? What the realist requires is an account of:

²⁰Wright himself makes this point, he notes that "nothing at all corresponds to the [ability to recognise whether the truth-condition obtains] in the case of statements for which the anti-realist finds the realist's account of understanding problematical" (Wright, 1993a, p. 17).

A practical ability which stands to understanding an evidence-transcendent truth-condition as recognitional skills stand to decidable truth-conditions. (Wright, 1993a, p. 23).

It is at this point that the manifestation argument ceases to be an argument which demonstrates that realism must be false, but instead becomes a challenge which the realist must overcome for their position to be tenable.²¹ As Wright notes, this argument doesn't constitute a "general, conclusive proof that [the realist position] cannot [be made coherent]" (Wright, 1989/1993b, p. 250); it doesn't prove the realist's position to be false. Instead, it merely represents a challenge to the realist, where a challenge is:

an argument that a number of beliefs, which are held individually to be attractive, are, if not outright inconsistent, at any rate in tension with each other. (Wright, 1989/1993b, p. 246).

The beliefs in tension are the first three premises of the argument. And the tension comes from the fact that **1** and **2** assert that understanding of declarative sentences consists in practical abilities that can be identified with grasp of their truth-conditions. The realist's problem is that they have provided us with no account of any abilities which can account for **3** — our ability which constitutes grasp of the truth-conditions of an undecidable sentence. And if no such account is available then that means that realism must be false.

1.4.4 The anti-realist solution

In contrast, the anti-realist has no such problem. They claim that **4** applies globally to all declarative sentences — our grasp of the truth-conditions of both decidable and undecidable sentences consists in our ability to recognise whether they are true or false — and respond to the reductio by rejecting **3** (evidence-transcendent truth).²²

²¹As I noted above, Byrne (2005) makes this change clear by drawing a distinction between the manifestation "argument" and the manifestation "challenge". In Miller's terminology (Miller, 2002), this is the distinction between the "strong" and "weak" versions of the argument.

²²Earlier anti-realist work used to respond to the reductio by rejecting **2** instead of **3**. For example, in "Truth" Dummett (1959) defines realism as the view that meaning essentially involves truth and falsity:

Realism consists in the belief that for any statement there must be something in virtue of which either it or its negation is true: it is only on the basis of this belief that we can justify the idea that truth and falsity play an essential rôle in the notion of meaning of a statement, that the general form of an explanation of meaning is a statement of the truth-conditions. (Dummett, 1959, p. 157).

And in "Realism" (Dummett, 1963/1978b) he explains the contrast between realism and anti-realist in terms of realists endorsing **2** and anti-realists rejecting **2**:

For the anti-realist, an understanding of such a statement consists in knowing what counts as evidence adequate for the assertion of the statement, and the truth of the statement can consist only in the existence of such evidence. For the realist, the notion of truth plays a more crucial rôle in the manner of determining the meaning of the statement. To know the

Anti-realists instead claim that “‘truth’ in ‘truth-conditions’ [denotes] an evidentially constrained notion” (Wright, 1989/1993b, p. 249), and that what it is to say that a sentence is true or false implicitly involves the notion of a warrant. For example, Wright does this by endorsing superassertibility, which he defines as follows:

A statement is superassertible . . . if and only if it is, or can be, warranted and some warrant for it would survive arbitrarily close scrutiny of its pedigree and arbitrarily extensive increments to or other forms of improvement of our information. (Wright, 1992, p. 48).

For potentially evidence-transcendent sentences, such as:

- James II suffered a migraine on the afternoon of his 32nd birthday.
- Goldbach’s conjecture: every even number greater than two is the sum of two primes.

it is clear that we are not currently in a position to be warranted in asserting or denying such sentences, so the anti-realist would simply refrain from saying that they are either true or false. As was pointed out in the section “Bivalence and realism” in §1.1.3, this would mean that the anti-realist would reject bivalence about such sentences. Rejecting bivalence does not mean that they think such sentences have a third truth-value, but merely that we are not justified in asserting that they must have either one or the other (i.e. **3** is false), since we don’t know if it would be ever possible to be warranted in asserting or denying such sentences.

Replacing **3**’s definition of truth with an evidence-constrained notion, such as superassertibility, clearly fits well with endorsing **4** for all declarative sentences; **1** and **2** together give us the claim that understanding is an ability which consists in the grasp of a sentence’s truth-conditions, and if truth is viewed as essentially evidence-constrained then it’s perfectly reasonable to identify understanding with the ability to recognise a sentence’s truth or falsity.

Also, because they reject **3** the anti-realist blocks the inference to **5** and the path to the reductio.

1.5 McDowell’s response: the argument from the truth-assertion platitudes

But what of the realist’s response to the challenge? Clearly they can’t agree that our understanding of a declarative sentence’s truth-conditions is constituted by the ability to

meaning of the statement is to know what it is for the statement to be true (Dummett, 1963/1978b, p. 155).

recognise its truth-value, since such a view seems forced to adopt an evidence-constrained view of truth. So they must provide an account of how our other abilities can constitute our grasp of truth-conditions, and show how our use of these demonstrates that we have a realist understanding of truth.

In “Anti-realism and the Epistemology of Understanding” McDowell (1981/1998b) makes the first steps towards a response to this anti-realist challenge. This paper was written back in 1981, before Wright’s refinements to the manifestation argument, and so is only intended to constitute a response to Dummett’s earlier version. However, I think it provides the best introduction to the general shape of McDowell’s response, and then later in the chapter I will explore how it can be refined to respond to Wright’s more sophisticated formulations of the argument.

McDowell explains his argument as follows

Knowledge of what a sentence can be used to assert is knowledge that can be directly manifested, on appropriate occasions, by using the sentence in such a way as manifestly to assert precisely that. Of course one may never have occasion for the assertoric utterance of some sentence that is within one’s syntactic reach; but no one, however anti-realistically inclined, should suppose this suffices to show that one does not know what one would be asserting if one did have occasion to utter the sentence assertorically. Specifications of contents of potential assertions are, by way of our platitude, specifications of conditions under which the sentences used to effect those assertions would be true. Now if a sentence lacks an effective decision procedure, then the condition that any competent speaker knows he would be asserting to obtain if he used the sentence in order to make an assertion — which in fact is a condition under which the sentence would be true, whether or not a theory of meaning explicitly calls it that — is *ex hypothesi* not a condition whose obtaining, if it does obtain, a competent speaker can be sure of being able to put himself in a position to recognize. Thus, without lapsing into psychologism,²³ we seem to have equipped ourselves with a kind of realism: a description of linguistic competence that makes central use of the idea that speakers have a knowledge of conditions which they are not, in general, capable of recognizing whenever they obtain. (McDowell, 1981/1998b, pp. 321–322, emphasis in original).

This argument is dense, so I’m going to break it down into the following six steps:

- i. “Knowledge of what a sentence can be used to assert is knowledge which can be directly manifested . . . by using the sentence . . . to assert precisely that.” (McDowell,

²³The avoidance of the threat of psychologism is a very significant part of McDowell’s paper. Psychologism is the view, held by people like Chomsky and Fodor, that the significance of someone else’s utterances can only be guessed at, and is hidden away in a private sphere. Both Dummett and McDowell agree that Wittgenstein has revealed psychologism to be fatally flawed, and that we must instead accept that there is nothing more to our understanding than what can be manifested publicly. Dummett’s arguments effectively aim to show that anti-realism is a consequence of this rejection of psychologism. McDowell’s goal in the paper is to show that we can still be realists whilst respecting Wittgenstein’s anti-psychologistic insight.

1981/1998b, p. 321).

For example, by using the sentence ‘The Eiffel Tower is in Paris’ to assert *that The Eiffel Tower is in Paris* we are able to manifest to other English speakers that we know that the sentence is used to assert *that The Eiffel Tower is in Paris*.²⁴

McDowell’s next step is to claim that:

Specifications of contents of potential assertions are, by way of our platitude, specifications of conditions under which the sentences used to effect those assertions would be true. (McDowell, 1981/1998b, pp. 321–322).

So, when we use the sentence ‘The Eiffel Tower is in Paris’ to assert *that The Eiffel Tower is in Paris* the content of that assertion is the conditions in which it would be true or false that The Eiffel Tower is in Paris — this is effectively **2**, the truth-conditional conception of understanding. If we combine **i** and **2** we get the conclusion that when making that assertion we are manifesting our knowledge of the conditions in which it would be true or false that The Eiffel Tower is in Paris. This gives us:

- ii. The knowledge a speaker manifests when they use a sentence to make an assertion is knowledge of the sentence’s truth-conditions.

This is just the truth-conditional conception of understanding, premise **2** of the manifestation argument, and, according to McDowell, a platitude.

Now if a sentence lacks an effective decision procedure, then the condition that any competent speaker knows he would be asserting to obtain if he used the sentence in order to make an assertion — which is in fact a condition under which the sentence would be true, whether or not a theory of meaning explicitly calls it that — is *ex hypothesi* not a condition whose obtaining, if it does obtain, a competent speaker can be sure of being able to put himself in a position to recognize. (McDowell, 1981/1998b, p. 322).

We can reconstruct the argument in this sentence by breaking it down into three stages, firstly the obvious fact that:

²⁴To avoid any potential misunderstanding, it’s crucial to emphasise that McDowell is only claiming that we *can* directly manifest our knowledge of what a sentence can be used to assert by using the sentence to make an assertion. He is not claiming that *every* case of someone saying a sentence will involve them manifesting their knowledge of what the sentence can be used to assert. For example, consider someone who mistakenly believes that The Eiffel Tower is in fact The Arc de Triomphe. When they speak the words ‘The Eiffel Tower is in Paris’, their doing so would not be a case of manifesting their knowledge of what the sentence can be used to assert, since they don’t actually know what the sentence is used to assert. Anyone who hears them saying this sentence may mistakenly draw the conclusion that they have used the sentence to make the assertion *that The Eiffel Tower is in Paris*, but that would be a mistake. But the essential point for McDowell here is that the possibility of making such mistakes does not undermine the fact that (in cases when mistakes aren’t made) people are able to manifest their understanding of what the sentence can be used to assert.

iii. Speakers are able to use undecidable sentences to make assertions.

For example, we have no evidence either way as to the truth of ‘James II suffered a migraine in 1665, on the afternoon of his 32nd birthday’ or for Goldbach’s conjecture, and no finite procedure which is guaranteed to verify or falsify such sentences. And yet that doesn’t mean that we aren’t able to use such sentences to make assertions.

The second stage is to point out that from **iii** and **i** it follows that:

iv. Speakers can manifest knowledge of what undecidable sentences can be used to assert.

For example, speakers are able to use ‘James II suffered a migraine in 1665, on the afternoon of his 32nd birthday’ to assert *that James II suffered a migraine in 1665, on the afternoon of his 32nd birthday*, and by doing so they can manifest that they know that the sentence can be used to assert *that James II suffered a migraine in 1665, on the afternoon of his 32nd birthday*.

And the final stage is to point out that from **iv** and **ii** we get:

v. Speakers can manifest knowledge of the truth-conditions of undecidable sentences.

Which leads us to the desired realist conclusion:

vi. “we seem to have equipped ourselves with a kind of realism: a description of linguistic competence that makes central use of the idea that speakers have a knowledge of conditions that they are not, in general, capable of recognizing whenever they obtain” (McDowell, 1981/1998b, p. 322).

In summary, McDowell’s argument runs as follows:

- i.** “Knowledge of what a sentence can be used to assert is knowledge which can be directly manifested . . . by using the sentence . . . to assert precisely that.” (McDowell, 1981/1998b, p. 321).
- ii.** The knowledge a speaker manifests when they use a sentence to make an assertion is knowledge of the sentence’s truth-conditions.
- iii.** Speakers are able to use undecidable sentences to make assertions.
- iv.** Speakers can manifest knowledge of what undecidable sentences can be used to assert.
- v.** Speakers can manifest knowledge of the truth-conditions of undecidable sentences.
- vi.** “we seem to have equipped ourselves with a kind of realism: a description of linguistic competence that makes central use of the idea that speakers have a knowledge of conditions that they are not, in general, capable of recognizing whenever they obtain” (McDowell, 1981/1998b, p. 322).

However, there are three problems with this argument, and I shall use the rest of this chapter to respond to them in turn. The problems are:

Problem 1: What justifies premise **i**? (That “Knowledge of what a sentence can be used to assert is knowledge which can be directly manifested . . . by using the sentence . . . to assert precisely that.” (McDowell, 1981/1998b, p. 321).)

Problem 2: How can McDowell respond to Wright’s refined version of Dummett’s manifestation argument, which accepts that understanding is constituted by grasp of a sentence’s truth-conditions, but instead argues that we should view truth as epistemically-constrained?

Problem 3: If this argument works wouldn’t it justify realism for all discourses that permit the formulation of undecidable sentences? If so, this would be problematic for McDowell because in “Mathematical Platonism and Dummettian Anti-Realism” (McDowell, 1989/1998g) he advocates anti-realism for mathematics, and mathematics clearly includes undecidable sentences (such as Goldbach’s conjecture).

1.6 Solution to problem 1: modesty

The first of our three problems is the question of what justifies the first premise:

i. “Knowledge of what a sentence can be used to assert is knowledge which can be directly manifested . . . by using the sentence . . . to assert precisely that.” (McDowell, 1981/1998b, p. 321).

What McDowell is effectively doing here is suggesting an alternative to **4** in Wright’s argument. Remember that Wright’s original argument didn’t show that **4** necessarily followed from **1** and **2**, it was merely a suggestion for an ability which fits the requirements of those two premises — being able to recognise the obtaining of a sentence’s truth-conditions is a publicly manifestable ability that constitutes grasp of a sentence’s truth-conditions. With **i** McDowell is effectively putting forward an alternative ability which also fits the requirements of **1** and **2** — a speaker can publicly manifest their grasp of the sentence’s truth-conditions simply by using that sentence to make an assertion.

One potential objection to this suggestion, and this is an objection which Dummett would raise (at least according to McDowell’s presentation of Dummett’s position in “Anti-realism and the Epistemology of Understanding” (McDowell, 1981/1998b)²⁵), is that merely asserting a sentence isn’t sufficient to publicly manifest your understanding in the way required by **1** (the Wittgensteinian conception of linguistic understanding).

²⁵In §7 (pp. 331–338) of that paper McDowell explores Dummett’s position on this, and criticises the epistemological assumptions which he thinks entice Dummett into supporting this view.

Remember that at the core of **1** was the claim that the abilities that constitute understanding must be publicly manifestable, so as to render them open to view to other speakers, so that they can be manifested in a way which is publicly evaluable as correct or incorrect. This is motivated by the claim that language is inherently public, and so meanings can't go beyond what can be manifested publicly. But consider a monolingual French speaker, if they heard someone say the sentence 'Snow is white' they may have no idea what was being said, and so the speaker would have failed to manifest to them their grasp of the sentence's truth-conditions.

The question here is whether the Wittgensteinian premise requires us to manifest our understanding to any speaker whatsoever, or merely to speakers of our own language. McDowell claims that part of learning a language involves acquiring "a perceptual capacity that is not necessarily universally shared" (McDowell, 1981/1998b, p. 332), which enables us to "hear more . . . after we have learned the language" (McDowell, 1981/1998b, p. 332). So, for example, our monolingual French speaker doesn't understand what an English speaker means by 'Snow is white', but if they were to learn English then when those words are spoken by an English speaker in English they would be able to hear that the speaker is saying that snow is white. In chapter 3 I'll evaluate in detail McDowell's arguments for this view, but for now I'll simply outline his position and assume that he's correct.

McDowell justifies the claim that language learning involves the development of these new perceptual abilities by pointing out that it is a requirement of modest theories of meaning. According to modesty the descriptions of the abilities that constitute sentential understanding in our theory of meaning must themselves involve intentional notions; we shouldn't expect our theories of meaning to reduce meaning to something non-intentional, but instead only give an account of meaning "from the inside". McDowell motivates this view by arguing that the alternative — Dummett's claim that theories of meaning must not assume understanding of intentional notions, in order to explain intentionality without presuming an understanding of it (explain intentionality "from the outside") — cannot escape devastating Kripkensteinian attacks which render any explanation of a sentence's meaning which doesn't use intentional notions hopelessly indeterminate.

As I said above, I'll evaluate this argument in chapter 3, but if we accept modesty, then the descriptions of the abilities that constitute sentential understanding must look something like this:

Someone understands 'Snow is white' to mean *Snow is white* if and only if they have the ability to use 'Snow is white' so as to be understood by speakers of English to be expressing the thought *that snow is white*.

Understanding this description of the ability requires an understanding of 'expressing the thought *that snow is white*'. And so clearly this description can only be understood by

an English speaker who is able to see the meaning of these words when they read them. Because of this, supporters of modest theories of meaning are committed to the claim that:

[the content of sentences is not] something “beneath” the words, to which we are to penetrate by stripping off the linguistic clothing; rather, as something present in the words — something capable of being heard or seen in the words by those who understand the language (McDowell, 1987/1998e, p. 99).

And this ability to hear or see the meaning in the words is the perceptual capacity which is learnt when learning a language.

At this stage it’s important to remember that McDowell is not claiming that whenever someone speaks, members of their speech community who hear them will *necessarily* be able to understand the thought they intend to express.²⁶ If someone believes that the word ‘snow’ means *paper* then when they utter the sentence ‘snow is white’ they will not be expressing the thought *that snow is white* but will instead be attempting to express the thought *that paper is white*. Anyone who heard them say this sentence may well make the mistake of thinking that they have expressed the thought *that snow is white*, but this would be a mistake. Key for McDowell here is the view that our ability to “hear someone else’s meaning in his words” (McDowell, 1984/1998o, p. 253) is fallible. But in good cases when all goes well, we are still able use a sentence to manifest knowledge of what the sentence can be used to assert.²⁷ For example, when someone uses the sentence ‘snow is white’ who correctly understands all the constituent words and intends to express the thought *that snow is white* then suitable members of their speech community will be able to hear them expressing the thought *that snow is white*.

In summary, the view which McDowell’s arguments for modesty support is the view that:

the truth-condition of a sentence (its content) is audible or visible — to those who understand the language — whenever the sentence is meaningfully uttered (McDowell, 1987/1998e, p. 100).

Which means that:

the thought (say) that some table-tops are square can be heard or seen in the words “Some table-tops are square”, by people who would be able to put their own minds into these words if they had occasion to do so (McDowell, 1987/1998e, p. 99).

²⁶We saw this point in the footnote in §1.5 involving the example of someone who mistakenly believes that The Eiffel Tower is the The Arc de Triomphe.

²⁷I’ll go into more detail of McDowell’s account of fallible epistemic abilities, in particular the ability to hear what someone means from their talk, in §1.7.3.

And so, according to McDowell, when making an assertion via the use of the sentences ‘Snow is white.’ or ‘James II suffered a migraine on the afternoon of his 32nd birthday.’ a speaker is able to manifest to fellow English speakers their knowledge of what these sentences can be used to assert, and their grasp of these sentences’ truth-conditions. And if this is true, then it’s clear that premise **i** is justified.

1.7 Problem 2: epistemically constrained truth

However, Wright argues that even if we accept **i** McDowell’s argument still fails to justify realism.²⁸ He accepts that McDowell’s platitudes show that someone who understands a statement has a conception of the state of affairs which is necessary and sufficient for its truth, but argues that it doesn’t show that they have a “conception of how such a state of affairs can obtain *undetectably*” (Wright, 1993a, p. 19). It is this latter claim which is required to justify the claim, essential to realism, that we understand what it is for a sentence to be undetectably true. According to Wright, the fact that someone who grasps the meaning of a statement grasps its truth-conditions “does nothing to justify the idea that the notion of truth which the reference therein to ‘truth-conditions’ invokes is the realist’s *objective* truth” (Wright, 1993a, p. 19). The problem here is that McDowell’s argument leaves ambiguous whether the truth-conditions which we grasp are potentially evidence-transcendent or evidence-constrained. Even though we understand what it would be like for them to detectably obtain, we may well not understand what it would be like for them to undetectably obtain, and therefore have no reason to conclude that we have a grasp of potentially evidence-transcendent truth-conditions, and so the argument fails as an attempt to demonstrate that our understanding is realist.

In effect, what Wright is claiming is that **ii** is ambiguous, and could be interpreted as either:

ii-AR. The knowledge a speaker manifests when they use a sentence to make an assertion is knowledge of the sentence’s *evidence-constrained* truth-conditions (i.e. an understanding of the conditions in which our evidence would justify asserting that the sentence is true or false).

or:

²⁸Like McDowell, Wright responds to Kripkensteinian arguments by endorsing non-reductionism about meaning, and he agrees with McDowell that we can’t give an account of meaning and intentionality “from outside”. It’s worth noting that (at least in the Introduction to *Realism, Meaning & Truth*) Wright doesn’t take an explicit stance on whether he agrees with McDowell’s claim that theories of meaning must be modest. Fortunately Wright’s argument in this section seems to be completely independent of the issues surrounding modesty; even if Wright were to completely agree with McDowell’s views on theories of meaning this argument would be unchanged.

ii-R. The knowledge a speaker manifests when they use a sentence to make an assertion is knowledge of the sentence’s potentially *evidence-transcendent* truth-conditions (i.e. an understanding of what it would be for the sentence to be true or false, even in cases where its truth-value is undecidable).²⁹

This ambiguity is problematic because McDowell’s realist conclusion is only justified if we read **ii** as **ii-R**, but McDowell’s platitudes do nothing to justify that reading over **ii-AR**.

This leads Wright to demand that McDowell³⁰ provide:

a description — appreciable by anyone who understands whatever kind of statement is giving rise to the problem — of what specifically, in the exercise of an understanding of such statements, manifests the fact that it consists in grasping a potentially evidence-transcendent truth-condition. (Wright, 1989/1993b, p. 254).

²⁹It’s important to note that this way of drawing the distinction between McDowell’s position and anti-realism is not quite in line with how he draws it in the later paper “Mathematical Platonism and Dummettian Anti-Realism” (McDowell, 1989/1998g). In that paper he clarifies his position as follows:

a realist (in the relevant sense) need not go beyond rejecting the distinctive thesis of anti-realism — the thesis that an account of how understanding connects with the world must restrict itself to decidable circumstances. (Anti-realism — verificationism, in a suitably broad sense — is the positive thesis here. If this is hard to remember, it may help to conceive the position I am talking about as anti-anti-realism rather than realism.) There is no need for an opponent of generalized anti-realism to volunteer a quasi-platonistic positive view, to the effect that we can know what it would be for some utterance to be true even though verification, or less than conclusive evidence, is not even in principle available. (Far better to leave to anti-realists the unattractive task of trying to make such uses of “in principle” clear.) (McDowell, 1989/1998g, pp. 355–356).

Crucially in this passage McDowell makes the claim that the form of realism (“anti-anti-realism”) which he’s defending is not committed to claiming that “we can know what it would be for some utterance to be true even though verification, or less than conclusive evidence, is not even in principle available” (McDowell, 1989/1998g, p. 356), instead it merely rejects the anti-realist’s claim “that an account of how understanding connects with the world must restrict itself to decidable circumstances” (McDowell, 1989/1998g, pp. 355–356).

This is significant because in the earlier paper “Anti-realism and the Epistemology of Understanding” he defines the realist position he’s defending as “a description of linguistic competence that makes central use of the idea that speakers have a knowledge of conditions that they are not, in general, capable of recognizing whenever they obtain” (McDowell, 1981/1998b, p. 322). So in the earlier paper he seems to be claiming that realism requires grasp of what it is for a sentence to be undetectably true, whilst in the latter paper he drops that requirement and instead merely claims that our understanding of what it is for a sentence to be true need not be restricted to cases where we can find evidence for its truth or falsity.

This means that my definition of **ii-R** isn’t quite in line with the form of realism McDowell defends in the latter paper, since **ii-R** involves the strong claim that speakers manifest knowledge of what it is for sentences to have undetectable truth-values which he goes on to drop. However, premise **ii** is meant as an interpretation of the earlier paper, and so I’ve decided to formulate **ii-R** and **ii-AR** along the lines of the realist / anti-realist distinction used by that paper. An exploration of the significance of this subtle shift in McDowell’s realism would be essential for a full understanding of his arguments here, particularly in the latter paper. However I don’t think it’s of great significance for my current concerns, and so for the rest of this chapter I shall ignore this detail to avoid unnecessary complexity.

³⁰In this particular passage Wright is in fact challenging Blackburn, rather than McDowell, to provide such a description, but the challenge is equally appropriate to McDowell.

McDowell's platitudes provide no such description, and so Wright concludes that McDowell has failed to justify his claim that our understanding of sentences involves them having potentially evidence-transcendent truth-conditions.

So, McDowell is left with the following question:

What could distinguish the performance of someone who had grasped [the possibility of some conditions obtaining undetectably] from that of someone who had not but who was prepared to count the statements as verified in the appropriate detectable circumstances? (Wright, 1980/1993c, p. 100).

And unless McDowell can provide us with a reason to believe that our grasp of the truth-conditions is realist then his argument from the truth-assertion platitudes has failed to respond to the challenge set by the manifestation argument.

However, it's important to emphasise that even without an argument to favour **ii-R** over **ii-AR** McDowell's argument still constitutes an improvement for the realist's position in the face of the manifestation argument. As we saw in §1.4.3 the manifestation argument ends up as a challenge for the realist to provide an account of:

A practical ability which stands to understanding an evidence-transcendent truth-condition as recognitional skills stand to decidable truth-conditions. (Wright, 1993a, p. 23).

At that point there didn't seem to be any candidates for such an ability. And so, although the manifestation argument doesn't prove realism to be false, it still revealed a very serious gap in the realist's position. If there is no manifestable ability that constitutes grasp of the realist truth-conditions of an undecidable sentence then it is clear that realism is false. And because no account of such an ability was on the table at that point, the anti-realist conclusion seemed inescapable.

Despite the ambiguity in **ii**, McDowell's argument from truth assertion platitudes does describe an ability which *could* constitute a realist grasp of the truth-conditions of an undecidable sentence and fill this gap in the realist's picture. The ambiguity leaves open whether or not our ability to use a sentence to make an assertion actually *does* constitute grasp of sentence's realist truth-conditions, and so it prevents us from justifiably concluding the realist to have won the argument. But at the very least we now have a candidate for an ability that may play that role, and so the realist's position is much less hopeless.

At this point we could get into a deep and subtle argument about which position is better off, on whom the burden of proof lies, etc. Fortunately McDowell has an argument in the later paper "Mathematical Platonism and Dummettian Anti-Realism" (McDowell, 1989/1998g) which can be used to remove the problematic ambiguity in **ii** and show that we should favour the **ii-R** reading, so there is no need to get bogged down in such messy problems.

1.7.1 Figuring in speech

However, before we get to McDowell’s argument in “Mathematical Platonism and Dummettian Anti-Realism” it’ll be useful to set the scene by looking at the exchange between him and Wright which led up to it. In “On “The Reality of the Past” ” McDowell (1978/1998k) runs the following argument for realism:

There is no real difficulty here. Acquiring one of the problematic conceptions is acquiring competence with the relevant part of a language. Exercising the conception, then, is nothing but exercising the relevant linguistic competence, in speech of one’s own or in understanding the speech of others. . . . The possibility of its figuring in his thoughts is secured . . . by the possibility of its figuring *in his speech*. A competent speaker has words to express, if need be, what state of affairs it is about whose perhaps undetectable obtaining he is capable of, for instance, self-consciously speculating, or understanding a fellow-speaker of his language to be speculating.

“When I imagine that someone who is laughing is really in pain I don’t imagine any pain-behaviour, for I see just the opposite. So *what* do I imagine?” — I have already said what.

(McDowell, 1978/1998k, p. 310 (the final quote is from *Philosophical Investigations* §393)).

McDowell’s argument here is that we can manifest our grasp of a potentially undetectable state of affairs simply by using the relevant words, nothing more is needed.

However, Wright responds in “Realism, Truth-value Links, Other Minds and the Past” (Wright, 1980/1993c, pp. 104–105) by claiming that this line of argument makes things far too easy for the realist. If this argument worked then it would be ridiculously trivial to save realism about mathematics:

Clearly, something has gone wrong here: for if this response to the manifestation-challenge were feasible, a classical realist about mathematics, for example, could meet the challenge simply by calling attention to the currency of classical logic. He could claim that we manifest our understanding of the possession of possibly proof-transcendent truth-conditions by Fermat’s ‘Last Theorem’ simply in regarding the law of Excluded Middle as validly applicable to it. (Wright, 1980/1993c, p. 104).

And in “Mathematical Platonism and Dummettian Anti-Realism” McDowell (1989/1998g) himself *rejects* realism about mathematics, so there must be something seriously wrong if his argument for realism elsewhere applies so trivially to mathematics.

Wright also points out that the correct interpretation of statements such as “ ‘Jones, for all that he gives no sign of it, may be in great pain’ ” (Wright, 1980/1993c, p. 104) is precisely what the anti-realist is questioning, and so the realist can’t simply presume their own interpretation to be correct. It’s not enough for the realist to simply point out that we usually interpret such sentences to have a realist meaning:

if we are realists and sincerely offer such expressions of our conviction, we can scarcely expect the mere fact that we can do so, intelligibly and correctly as it seems to us, to carry the day. For the fact is that it is possible to hold a mistaken philosophical theory about the character of our understanding of a particular class of statements, and for that theory to insinuate itself into our linguistic practices, motivating speculations and claims which we might otherwise not be prepared to make. (Wright, 1980/1993c, p. 105).

Why, Wright asks, can't we instead interpret the remark about Jones' pain as merely a way of expressing that statements about others' mental states are defeasible. When we say that Jones, despite his outward behaviour, may actually be in great pain we are not manifesting our grasp of what it is for Jones' pain to be undetectable, but simply reminding ourselves that any statement about someone's mental states is always potentially subject to correction when we are presented with new evidence. This reading is perfectly compatible with the anti-realist's claim that we don't grasp what it would be for Jones to be in undetectable pain and that we should endorse **ii-AR** instead of **ii-R**.

And when considering whether we should adopt realist or anti-realist interpretations of such sentences (effectively whether we should endorse **ii-R** or **ii-AR**) he argues that:

All of these would be possibilities which a Martian visitor, attempting radically to interpret our linguistic practice, would have to consider. And the manifestation-challenge will be met only when one of the two things is achieved: *either* we must call attention to features of our practice which would provide the Martian with a sound motive for preferring to attribute to us at least a bare realist understanding of statements of the relevant sorts, rather than advance any of the alternative accounts; *or* we must divine, a priori, pure philosophical considerations which constrain the former interpreting of behaviour which, of itself, fits the alternative accounts. (Wright, 1980/1993c, pp. 105–106, emphasis in original).

So, according to Wright, McDowell needs to provide additional evidence to defend his position. This evidence must take one of two forms: either it be such that it would convince a Martian who didn't already speak our language and was deciding how to interpret our talk, or it must take the form of an *a priori* philosophical argument. Also if it is to remain compatible with McDowell's position in "Mathematical Platonism and Dummettian Anti-Realism", whatever type of evidence he provides must not be equally applicable to mathematics in demonstrating the truth of mathematical realism.

1.7.2 McDowell's argument for anti-realism about mathematics

As I indicated above, despite being a realist about the vast majority of discourses, in "Mathematical Platonism and Dummettian Anti-Realism" McDowell (1989/1998g) endorses anti-realism about mathematics. At the core of McDowell (and Dummett and Wright's — they run the same argument; this is in effect a re-telling of the manifestation

argument) argument for anti-realism about mathematics is the Wittgensteinian conception of linguistic understanding which forms the premise **1** of the manifestation argument. This involves the claim that there is nothing more to our understanding than what can be manifested publicly — that “Meaning cannot transcend use” (McDowell, 1981/1998b, p. 344). According to the anti-realist argument, realism about mathematics violates this restriction.

Consider Goldbach’s conjecture, which is that every even number greater than two is the sum of two primes. According to realism about mathematics:

the meaning of the conjecture [is associated with] a specific possible configuration in arithmetical reality (the state of affairs that would obtain if the conjecture were true); and that configuration must determinately either obtain or not obtain, regardless of our lack of an effective procedure for deciding which of these alternatives is the case. (McDowell, 1989/1998g, p. 345).

McDowell’s explanation of the argument against such a view goes as follows:

The [mathematical realist] ... has it that our understanding of the relevant sentences consists in a conception of what it would be for them to be true that might outrun what we could display in sensitivity to proofs and refutations — even including all those that could in principle be devised. But what other aspect of our use of arithmetical sentences could be cited as potentially manifesting the supposed residue of understanding? There seems to be no acceptable answer to this question. Affirmation of the principle of bivalence without restriction, or conformity with classical methods of inference will not serve; the [mathematical realist] wanted these practices to be *justified* by an *independently* acceptable account of the relation between meaning and truth. So [mathematical realism] apparently makes no concession to the thought that someone’s understanding of a sentence must be able to be made fully overt in his use of it. The trouble is that this leaves it a mystery how one person can fully know another person’s meaning. (McDowell, 1989/1998g, p. 347, emphasis in original).

Central to the argument here is the claim that in mathematics there is nothing in our talk which can manifest our understanding of mathematical sentences other than our ability to construct proofs and refutations; as he says earlier in the paper “There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations.” (McDowell, 1989/1998g, p. 354). Because Goldbach’s conjecture is undecidable this means that the mathematical realist’s claim to have a conception of what it is for Goldbach’s conjecture to be true or false conflicts with the Wittgensteinian conception of linguistic understanding. If we lack the ability to verify that a mathematical sentence’s truth-conditions obtain (or not) then there is nothing left in our use which can manifest an understanding of what it is for the sentence to be true or false, and so we can’t grasp what it would be for such a sentences to be true or false undetectably.

According to McDowell the problem with mathematical realism is that it involves a conception of what it is for sentences to be true “to which our ways of deciding whether they are true are simply irrelevant” (McDowell, 1989/1998g, p. 349). Instead, “the platonist proceeds as if it makes sense to suppose that it may just *happen* to be the case that *every* even number is the sum of two primes” (McDowell, 1989/1998g, p. 345, emphasis in original). This means that mathematical realism violates the Wittgensteinian conception of linguistic understanding, because it involves “[refusing] to tailor the meaning of mathematical statements to the use that we make of them” (McDowell, 1989/1998g, p. 350).

Because of this McDowell concludes that we should refuse to “distinguish what it would be for an arithmetical statement to be true from what it would be for a proof of it to be in principle feasible” (McDowell, 1989/1998g, p. 347), and he instead holds that “mathematical reality . . . has no properties beyond those it can be proved to have” (McDowell, 1989/1998g, pp. 360–361).

As we saw earlier in the chapter, Dummett and Wright claim that an argument like this one can be run for all discourses, and not just mathematics. But McDowell (1989/1998g, p. 357) argues that there is a significant difference between the epistemology of mathematics and that of other discourses, and that the global anti-realist conclusion is a result of mistakenly modelling the epistemology of other discourses on that of mathematics. The reason the anti-realist argument works in mathematics is “because of the quite special role of verifications (proofs and refutations) *in the competent use of mathematical sentences in particular*” (McDowell, 1989/1998g, p. 354, emphasis in original), but the epistemological talk of other discourses is richer, and so the anti-realist argument is not applicable.

1.7.3 Wright’s epistemological mistake and McDowell’s fallibilism

In order to get clear on why McDowell thinks that the epistemological practice involved in most discourses is richer than that of mathematics it’ll be useful to start by explaining the epistemological picture of mathematics that Wright and McDowell agree upon, before then going on to explain Wright’s view of the epistemology of the natural world and what McDowell thinks is wrong with it. But before I start it’s important to note that the debate between epistemological positions like Wright’s and McDowell’s is still very much a live debate in epistemology, and to go into it in any real depth would involve moving far beyond the core focus of this chapter. So instead I intend to use this section to merely give the reader a rough picture of the debate as McDowell presents it in “Mathematical Platonism and Dummettian Anti-Realism”.³¹

Consider Pythagoras’ Theorem — that for a right angled triangle the square of the

³¹For a more thorough explanation of McDowell’s arguments for fallibilist epistemology, and a clear account of his general epistemic view see *Perception as a Capacity for Knowledge* (McDowell, 2011).

hypotenuse is equal to the sum of the squares of the other two sides — this is a true mathematical theorem which is justified by a proof. It would be fair to say that, at least for a mathematician who understands it, the proof warrants belief in Pythagoras' Theorem. In addition, the warrant the proof provides is indefeasible: the proof entails the truth of Pythagoras' Theorem. If we were to, somehow, discover that Pythagoras' Theorem was false this would mean not only that there was a problem with the proof, but that we didn't even have a proof in the first place.

Now consider a case involving the epistemology of the natural world. Wright's view of what is required for a belief to be verified is as follows:

S is capable of *actual verification* if and only if there is some investigative procedure, *I*, such that (i) we can actually implement *I* and, on that basis, achieve, if we are rational, what we will consider to be grounded belief in *S*; but (ii) subsequent grounds sufficient to call *S* into question would, if we are rational, have to be allowed to call into question simultaneously the truth of at least one member of the original *I*-class. (An undecided *S* may thus be regarded as *decidable in practice* just in case we have recognized that either it or its negation is capable of actual verification.) (Wright, 1982, p. 217, emphasis in original).

Consider the case in which you're looking at a snooker ball and the ball appears to be red. According to Wright the belief that you are looking at a red snooker ball would be warranted by the combination of the belief that you're in conditions which are appropriate for having accurate colour vision — such as that the light conditions are normal, you're not overly tired, you're not wearing tinted glasses, etc. — and the fact that it seems to you that there's a red snooker ball in front of you. The crucial aspect of Wright's view here is that, as in the mathematical case, if you were to discover that the ball was actually white then you *must* infer that one of your warranting beliefs were false (for example, you may have actually been looking at it under a red light which made it look red, and so the lighting conditions were not actually normal). If the falsity of your conclusion doesn't imply that there was something wrong with your warranting beliefs then that means that those beliefs are not sufficient for justifying belief in the conclusion, or for giving you knowledge in the case when the lighting conditions, etc., are good.³²

McDowell argues that the problem with this account of what is required for a belief to be verified is that it ends up making knowledge in most domains virtually impossible (or at least it would make knowledge only possible in cases where we have followed “a procedure that would strike anyone uncorrupted by philosophy as neurotic” (McDowell,

³²One difference between Wright's view of the mathematical case and his view of the case of knowledge about the natural world is that the warrant for the belief that there's a red ball is defeasible — even in the case where the ball is actually white it would still be fair to say that your false belief was warranted, whereas in the mathematical case if Pythagoras' Theorem turned out to be false then you can't have even had a proof in the first place.

1989/1998g, p. 360)). Consider the sentence ‘That is a tomato’. McDowell argues that Wright’s view cannot account for how we are able to verify the truth of that sentence by “the ordinary procedure of looking and seeing” (McDowell, 1989/1998g, p. 360), since even in ideal conditions for viewing tomatoes the belief that that is a tomato, arrived at by looking “may be wrong *simply* because what looks like a tomato is in fact a cunningly hollowed façade” (McDowell, 1989/1998g, p. 360). In such a case there is nothing wrong with any of the beliefs about our environment that justified our conclusion, we are in normal light, are not tired, etc., but still, the conclusion is false. According to Wright’s account if the falsity of the conclusion doesn’t imply that there’s something wrong with these warranting beliefs then that means that our investigative procedure isn’t rigorous enough to ever give us knowledge. So this would mean that even in cases where there is a tomato being in those conditions alone can’t be sufficient for knowledge that there is a tomato. But what more could be required? Does this mean that looking and seeing isn’t sufficient for knowledge that there is a tomato in front of us; does knowledge require us to cut the tomato open and check what’s inside? But then what about the possibility that it is in fact a different species of fruit which to the casual observer looks very much like a tomato, must we consult a botanist every time we need to know that there is a tomato in front of us? This is clearly ridiculous, and so there must be something deeply wrong with Wright’s epistemological picture.

According to McDowell, Wright’s mistake is to fail to recognise “that the substantiality — the non-flatness — of some empirically accessible states of affairs carries with it a *quite additional* fallibility in our cognitive access to them, something with no counterparts in mathematics” (McDowell, 1989/1998g, p. 362, emphasis in original). McDowell thinks that Wright is correct to think that in mathematics if our conclusion is wrong then there must be something wrong with the proof which grounded that belief — if Pythagoras’ Theorem turned out to be false then there *must* be something wrong with our apparent proof for its truth. And it would be wrong to use a procedure in mathematics which, when executed correctly, wasn’t guaranteed to produce correct results. Wright’s mistake is to attempt to generalise this principle to the epistemology of other discourses, such as the natural world. McDowell holds that in the case of mathematics the only possible sources of error reside in our correct execution of the investigative procedure, whereas in the epistemology of the natural world we must also take into account the “*sheer misleadingness in the objects* on which an investigative procedure is exercised, as in the case of the tomato façade” (McDowell, 1989/1998g, p. 363, emphasis in original). Justified beliefs about the natural world don’t require us to follow a procedure so rigorous that the falsity of our conclusion implies that we must have executed it wrongly — in the case where we’re looking at an actual tomato we don’t need to break the tomato open to check it’s not a façade in order to be justified in believing that it is a tomato. Instead, what we need to do is:

to see our way to taking seriously the idea that experience is an essentially fallible, but unmediated, openness on our part to a reality that includes substantial (non-flat) states of affairs (McDowell, 1989/1998g, p. 362).

According to McDowell’s epistemological picture we have a fallible ability which, when it goes right (such as when we’re looking at a tomato in good light), enables knowledge, but it can also sometimes go wrong and lead to false beliefs (such as in the case when what we’re looking at is actually a tomato façade). When it does go wrong that’s not necessarily because there’s anything wrong with our execution of the investigative procedure — we could be in perceptual conditions which are perfectly good for enabling knowledge when things go right (being in good light, not being tired, not wearing tinted glasses, etc.) — it’s just that the ability to know by looking is itself fallible, and sometimes goes wrong.³³

1.7.4 The link between fallibilism and realism

Now that I’ve got clear on McDowell’s view of how the epistemology of mathematics is different from that of the natural world I can explain how it connects up to the realism / anti-realism debate and what’s special about “the quite specific role of verifications (proof and refutations)” (McDowell, 1989/1998g, p. 354) in mathematics.

The Wittgensteinian conception of meaning and understanding that is at the heart of these debates claims that “Meaning cannot transcend use” (McDowell, 1981/1998b, p. 344), and so our conception of what it is for sentences to be true or false (an aspect of their meaning) must fit with our ways of discovering their truth-value (an aspect of our use). But McDowell claims that “There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations.” (McDowell, 1989/1998g, p. 354). It is here that the significance of Wright’s epistemological picture becomes clear. The only scope that picture has for accounting for the possibility of making

³³I should note at this point that this presentation of McDowell’s epistemology leaves out some details which are essential for a complete understanding of his view, but would only serve to overcomplicate the current dispute with Wright.

The key aspect of McDowell’s view which isn’t mentioned above is that he thinks that “The warrant for belief that the [perceptual state of seeing one’s environment] provides is indefeasible; it *cannot* be undermined.” (McDowell, 2011, p. 31). McDowell argues that in a good epistemological case, such as when looking at a real tomato, “a perceptual state can consist in a subject’s having a feature of her environment perceptually present to her” (McDowell, 2011, p. 31), and if the subject’s experience actually makes the tomato present to her it is *impossible* that she may actually be looking at a mere façade. It’s true that in the bad case, when the subject is looking at a mere tomato façade, she may be unable to distinguish the experience from the good case — “one can mistake a defective exercise of the capacity for a non-defective exercise” (McDowell, 2011, p. 42) —, but McDowell argues that despite this the two experiences are different in an epistemologically significant sense. When the subject’s perceptual capacity works correctly it involves “perceptual states that make features of the environment present to them, and so provide conclusive warrant for corresponding beliefs” (McDowell, 2011, p. 34). The fallibilism in this view consists in the fact that our perceptual capacity itself is fallible, in that it “does not ensure that its possessor is always in a position to discriminate defective exercises from non-defective exercises” (McDowell, 2011, p. 39).

mistakes — the possibility of there being truths which we don't currently have knowledge of — is in the case of where we've made a mistake in our reasoning. All such mistakes are due to us failing to execute our investigative procedure correctly, and make no use of the idea that there could be facts which are undetectable. If there's nothing to our mathematical practice other than the constructing of proofs and refutations, then there is no space for the claim that the discourse involves grasp of undetectable truths; no reason to believe that mathematical reality “has [any] properties beyond those it can be proved to have” (McDowell, 1989/1998g, pp. 360–361).

In contrast McDowell's epistemic fallibilism involves the idea that we can make a mistake in cases where even when we've executed our investigative procedure completely and correctly we can't *guarantee* that we've come to the right conclusion. For discourses where a fallibilist epistemology is appropriate speakers must be able to make sense of the idea that that thing which looks like a tomato, which we are looking at in conditions which can be sufficient for knowledge, may actually be a mere façade. The epistemological picture essentially involves the view that our capacity to perceive the world is *itself* fallible and that we get things wrong because of a “*sheer misleadingness in the objects*” (McDowell, 1989/1998g, p. 363, emphasis in original). And so for those discourses there is evidence in our practice which justifies the claim that we have a realist understanding of truth, and that “we can contemplate in thought a reality that is determinate *beyond our access to it*” (McDowell, 1989/1998g, p. 355, emphasis in original).

1.7.5 The solution to problem 2

Wright's original criticism of McDowell's argument for realism from truth assertion platitudes was that premise **ii** was ambiguous, Wright effectively argued that we could read it as either:

ii-AR. The knowledge a speaker manifests when they use a sentence to make an assertion is knowledge of the sentence's *evidence-constrained* truth-conditions (i.e. an understanding of the conditions in which our evidence would justify asserting that the sentence is true or false).

or:

ii-R. The knowledge a speaker manifests when they use a sentence to make an assertion is knowledge of the sentence's potentially *evidence-transcendent* truth-conditions (i.e. an understanding of what it would be for the sentence to be true or false, even in cases where its truth-value is undecidable).

For McDowell's argument to work we must read **ii** as **ii-R**, but Wright claimed that McDowell had failed to provide any justification for preferring that reading over the

alternative. Wright pointed out that we can't simply refer to the fact that we make claims like “ ‘Jones, for all that he gives no sign of it, may be in great pain’ ” (Wright, 1980/1993c, p. 104) to justify the claim that we have a realist understanding of truth, since such claims can instead be interpreted as simply expressing the fact that statements about others' mental states are defeasible. This means that such claims are compatible with **ii-AR** and cannot be used to justify reading **ii** as **ii-R**. And he challenged McDowell to provide us with a reason to read **ii** as **ii-R** rather than **ii-AR**.

McDowell's above epistemological argument provides the required reason for reading **ii** as **ii-R**. As we've just seen from the arguments involving the tomato façade, for most discourses (including talk about the natural world) our very epistemology essentially involves the idea that the world can have properties beyond those we know of, or at least beyond those which are decidable. If a fallibilist epistemology is appropriate for a particular discourse then that means that in that discourse “we can contemplate in thought a reality that is determinate *beyond our access to it*” (McDowell, 1989/1998g, p. 355, emphasis in original). This shows that the notion of truth involved in that discourse must be realist, and that we understand sentences to have potentially evidence-transcendent truth-conditions — **ii-R** is true and not **ii-AR**.

This shows that McDowell's argument from truth assertion platitudes alone, as it stands in “Anti-Realism and the Epistemology of Understanding”, is not sufficient to demonstrate realism. We must add to it justification for viewing truth as potentially evidence-transcendent in each particular discourse where the argument is applied. As we've just seen, one way we can argue for a realist understanding of truth in a discourse is by showing that we should adopt a fallibilist account of the epistemology of that discourse. And so, for McDowell to justify belief in realism about a particular discourse, he can't rely on the platitudes cited in “Anti-Realism and the Epistemology of Understanding” alone, but must also make use of his epistemological framework.

1.8 Problem 3: mathematical realism

The solution to problem 2 appears to make the solution to problem 3 incredibly straightforward. Problem 3 was the worry that the argument from truth assertion platitudes in “Anti-realism and the Epistemology of Understanding” seems to apply generally to all discourses which permit the formulation of undecidable sentences. Mathematics is one such discourse, and so that argument should justify realism about mathematics, and yet McDowell explicitly advocates mathematical anti-realism in “Mathematical Platonism and Dummettian Anti-Realism”. We've now seen that the original argument alone isn't sufficient for demonstrating realism, and that further facts must be added to support the required reading of premise **ii**. According to McDowell these further facts are lacking in the case of mathematics — in that domain there is no justification for viewing the

notion of truth involved as realist — and so McDowell’s endorsement of mathematical anti-realism in the later paper seems perfectly legitimate.

However, it’s not quite that simple. In this section I’m going to argue that in actual fact McDowell and Wright’s account of mathematical practice is deeply flawed and based on a radical oversimplification of what mathematicians actually do. Once we correct this account of mathematics then it becomes clear that McDowell’s arguments actually do justify realism about mathematics too.

McDowell’s argument for an anti-realist view of truth in mathematics goes something like this:

- a. If we have produced a proof or a counterexample of a particular arithmetical claim then, if we haven’t made a mistake in our reasoning, our conclusion is guaranteed to be right.
- b. So, a fallibilist epistemology of mathematics would be inappropriate; any false conclusions we draw are a result of a mistake in *our* reasoning, and not a result of any misleadingness in the mathematical objects (if there are such things) themselves.
- c. So, “There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations.” (McDowell, 1989/1998g, p. 354).
- d. Conclusion: the notion of truth involved in mathematical talk is an anti-realist one; we have no understanding of what it is for a sentence to be true or false beyond having a proof or a counter-example (so we should read **ii** as **ii-AR**).

The problem with this argument is that **a** and **b** are not sufficient for justifying **c**, and that, in actual fact, **c** is false.

1.8.1 The problem with **c**

The problem with **c** (McDowell’s claim that “There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations.” (McDowell, 1989/1998g, p. 354)) is that there are in fact many ways to manifest linguistic competence in mathematics other than responses to proofs and refutations. This is significant because much of this other linguistic behaviour doesn’t seem to fit the epistemic picture which is appropriate for proofs and refutations, and so could have consequences as to whether we should be realists or anti-realists about mathematics. I’m going to focus on just one way mathematics goes beyond proofs and refutations, that of the assessment of the plausibility of conjectures. Here’s the philosopher of mathematics David Cornfield’s account of the role the notion of plausibility plays in mathematics:

mathematicians spend much of their time formulating conjectures, providing evidence for these conjectures by way of calculating particular instances,

or drawing analogies to similar problems, and then planning proofs based on the existing knowledge and their understanding of the proofs of related results. Much of this activity has an inductive flavour to it. . . . considerations of plausibility govern the choices of mathematicians at different time-scales, from their day-to-day reasoning on a specific problem to the decision to dedicate a considerable part of their lives to a particular research programme.

Considerations of which results are likely to be correct and which are likely to be accessible using available proof techniques are clearly relevant to rational decision making as to the choice of research topic. Furthermore, it is common practice to build a considerable body of mathematics upon unproven conjectures. For example, we have results which start ‘If the Riemann hypothesis is correct, then . . .’ (Cornfield, 2003, pp. 26–27).

The work involving the assessment of the plausibility of conjectures in mathematics as described here clearly goes far beyond mere proofs and refutations. Obviously the best way to support a mathematical conjecture is to prove it, and that is the final goal for any conjecture, but Cornfield talks of many other ways mathematicians provide evidence for conjectures, such as calculating particular instances and by drawing analogies with similar problems. So, for example, for Goldbach’s conjecture even though we do not currently know a proof we may be able to provide evidence for its truth by testing particular cases and by showing how it is similar to other things that we have proved.

Cornfield also talks about the planning of proofs, which is clearly something that is often done when we currently lack a successful proof, and is a practice where we lack the kind of epistemic guarantees that we have once we’ve finished developing a proof. The planning and construction of a proof relies heavily upon intuition and speculation — a mathematician will try various techniques to prove their conjecture, and the more skilled the mathematician, the better they will be at guessing which technique is likely to provide a proof for a particular conjecture. So it seems that the construction of proofs itself requires skills far beyond those a philosopher may initially consider when thinking about the nature of proof in mathematics.

In addition, the fact that “it is common practice to build a considerable body of mathematics upon unproven conjectures” shows that mathematicians take considerations of plausibility very seriously. When a mathematician gives reasons why a conjecture is likely to be true they have no proof for that claim, and cannot be certain that the conjecture is one which will end up being helpful for the discovery of more truths. It seems unquestionable that being able to assess which conjectures to take seriously is another essential skill in competence in mathematics, and so **c** must be false.

To flesh this out, I’m going to use an example of a specific mathematician talking about the work involved in some of their own research:

I began studying 3-dimensional manifolds and their relationship to hyperbolic geometry. (Again, it matters little if you know what this is about.) I gradually built up over a number of years a certain intuition for hyperbolic three-manifolds, with a repertoire of constructions, examples, and proofs. . . . After a while, I conjectured or speculated that all three-manifolds have a certain geometric structure; this conjecture eventually became known as the geometrization conjecture. About two or three years later, I proved the geometrization theorem for Haken manifolds. It was a hard theorem, and I spent a tremendous amount of effort thinking about it. (Thurston, 1994/2006, p. 51).

What's stunning in this passage is that it took Thurston over two years to successfully develop the proof he was looking for. Up to then he describes his work as building up "intuition", intuition which no doubt guided him in his development of the proof, but which has no place in McDowell's picture of mathematical practice.

He then goes on to talk about his thoughts on the future of the field:

This particular proof probably has only temporary logical value, although it has a high motivational value in helping support a certain vision for the structure of 3-manifolds. The full geometrization conjecture is still a conjecture. It has been proven for many cases, and is supported by a great deal of computer evidence as well, but it has not been proven in generality. *I am convinced that the general proof will be discovered*; I hope before too many more years. (Thurston, 1994/2006, p. 54, emphasis added).

Here he talks about the "full geometrization conjecture" and claims to be "convinced that the general proof will be discovered". He lacks a proof for this claim, and yet he is convinced it is true. How could such conviction possibly be justified if, as McDowell claims, "There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations." (McDowell, 1989/1998g, p. 354)?

1.8.2 The significance of the falsity of **c**

In the previous section I explored a little of the practice of mathematicians in the face of conjectures, cases where we, by definition, lack a proof or a refutation. Yet, contrary to McDowell's claim, there seems to be a great deal of mathematical work being done here, work which seems essential for competence in mathematics. In the face of conjectures mathematicians are able to assess justification for belief in the conjecture, this work guides their entire practice and must be included in any full account of the epistemology of mathematics. It seems hard to deny that **c** is false.

A much more detailed analysis of mathematical practice needs to be done to work out the full significance of this mistake for McDowell and for the realism / anti-realism debate in mathematics in general. But to round off this chapter I'm going to offer a few brief notes on what I think the significance of how mathematicians work with conjectures may be.

The mathematical practice surrounding conjectures is interesting because, contrary to McDowell, Wright and Dummett's views on the epistemology of mathematics, it appears to involve more of a fallibilist epistemology, rather than one where all mistakes are due to our failing to follow though step-by-step logical processes correctly. When a mathematician evaluates the evidence for a particular conjecture this evidence can be so strong that it *convinc*es them to believe the conjecture to be true, and yet, since they don't yet have a proof, it is still possible that the conjecture is wrong. It seems at least possible that some of these firmly held conjectures may turn out to be false. But, like the case of the tomato façade, it doesn't seem that this means we must conclude that there was anything wrong with the investigative procedure which led us to believe in the conjecture. Cornfield pointed out that belief in conjectures may be supported by testing particular cases and by drawing analogies to things we already have a proof for. And yet when a conjecture justified in this way turns out to be false it seems that we must not necessarily reject either of these original grounds for believing the conjecture. If it turned out that Goldbach's conjecture was false it would not necessarily mean that when we'd tested particular cases, or when drawing analogies to similar cases which we had proved, we'd made a mistake. Instead, perhaps we should conclude that the mathematicians' ability to justify the truth of conjectures is itself fallible.

If this is the case then we seem to have a significant aspect of mathematical practice which essentially involves the realist's notion of truth. We may never discover a proof for Thurston's full geometrization conjecture, but that doesn't stop Thurston's talk essentially involving an understanding of what it would be for it to be true or false, since he still understands what it is to provide evidence for its truth, and for the facts to be different to what the evidence points to. And so here we have what looks like evidence that in mathematics we should read **ii** in the argument from truth assertion platitudes as **ii-R**, and so be realists about not only the natural world, but mathematics too.

1.9 Conclusion

I have identified three problems with McDowell's response to the manifestation challenge in "Anti-realism and the Epistemology of Understanding" (McDowell, 1981/1998b), but argued that all could be responded to by drawing on his other, later work. The first problem was McDowell's justification for the first premise, his claim that "Knowledge of what a sentence can be used to assert is knowledge which can be directly manifested . . . by using the sentence . . . to assert precisely that." (McDowell, 1981/1998b, p. 321). In response to this problem I drew on McDowell's work on modest theory of meaning in "In Defence of Modesty" (McDowell, 1987/1998e), and demonstrated that if McDowell is correct there then he is right in claiming that we can manifest our understanding of sentences to members of our speech community simply by speaking them.

I next noted that there is an ambiguity in the second premise of McDowell's argument. The premise:

- ii.** The knowledge a speaker manifests when they use a sentence to make an assertion is knowledge of the sentence's truth-conditions

is ambiguous because it doesn't make any claims about whether the truth-conditions involved are evidence-constrained or potentially evidence-transcendent. If we only manifest our knowledge of the sentence's evidence-constrained truth-conditions then the argument fails to justify realism. In response to this problem I drew on McDowell's work in "Mathematical Platonism and Dummettian Anti-Realism" (McDowell, 1989/1998g) and showed that if a fallibilist epistemology is appropriate for a particular discourse then it must be true that "we can contemplate in thought a reality that is determinate *beyond our access to it*" (McDowell, 1989/1998g, p. 355, emphasis in original). This justifies the realist reading of **ii** for that particular discourse because it shows that the notion of truth involved must be realist, and that we understand sentences to have potentially evidence-transcendent truth-conditions. An interesting result of this was that it shows that McDowell's response to the manifestation argument is discourse specific: it only justifies realism in discourses where we can give reasons for endorsing the realist reading of **ii**.

In "Mathematical Platonism and Dummettian Anti-Realism" McDowell not only argues for fallibilism about the natural world, he also endorses anti-realism about mathematics. I argued that the reason he does so is because he falsely believes that "There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations." (McDowell, 1989/1998g, p. 354). I demonstrated that this is wrong because mathematical competence requires epistemic practices in which mathematicians can have good reasons for making claims for which they have no proof, and they even go so far as justifiably believing that certain conjectures are true without having yet developed a proof for them. This justifies endorsing a fallibilist account of the epistemic practices of mathematicians, and so using my argument in response to the ambiguity in **ii** I concluded that we should be realists about mathematics.

Chapter 2

Rule-following part 1: The problem of rule-following

Over the next two chapters I shall look at Wittgenstein's rule-following problem. This problem focuses on the question of how it is possible for meanings, or more generally, rules, to come to mind. Essential to this problem is the fact that meanings are normative — there are right and wrong ways to use language and we can sometimes make mistakes in our applications of language: what seems right to us may not actually be right. McDowell's response to this problem plays a key role in his general philosophical picture and in the second of this pair of chapters I shall look at his position, and evaluate its success.

Although the foundations for this debate were laid in Wittgenstein's *Philosophical Investigations* (Wittgenstein, 1953/2001), the contemporary debate, including McDowell's response to the problem, has been heavily influenced by Kripke's work in *Wittgenstein on Rules and Private Language* (Kripke, 1982). I'm going to start by looking at Kripke's account of the problem, which is useful in demonstrating the argument at the heart of the problem, and in giving us a number of conditions that any direct solution to the problem must satisfy. In short, Kripke shows that any attempt to claim that there are facts in virtue of which we have meanings in mind must be able to explain how these facts can determine the extension of our expressions, and must explain the normativity of meaning — how is it that these facts can tell us how we *should* use language. Kripke himself responds to this problem by arguing that there are no facts which constitute meaning, but he attempts to avoid the apparent disastrous consequence that all language is meaningless by turning instead to assertion conditions. However, as Wright shows, this response fails, so I shall next turn to Wright's own work on rule-following.

Because much of McDowell's work on rule-following has been a response to that of Wright, I shall be looking at two different positions on meaning which Wright has advanced. The first is that in his book *Wittgenstein on the Foundations of Mathematics* (Wright, 1980), where he attempts to develop a communitarian account according to

which the norms governing a speaker's use of an expression are determined by how the community he belongs to is disposed to use it.¹ However, as Boghossian shows, Wright's early solution fails because it can't account for the idea that the community can sometimes have systematic dispositions to make mistakes.

The final response to the rule-following problem I shall look at in this chapter is Wright's judgement-dependent account of meaning. According to this view, what constitutes having a particular meaning in mind is the fact that in best conditions you would judge that that is what you mean. Crucially, this account involves the claim that the reason Kripke's argument initially seems so compelling is because of the implicit assumption that knowledge of meaning cannot be non-inferential. He argues that Kripke has no justification for this assumption, and that once we remove it we open the door to finding facts which constitute meaning. McDowell effectively makes the same claim, but from there his and Wright's accounts substantially differ. Wright claims that we need to develop a philosophical account of how there can be facts about meaning which we are able to have non-inferential knowledge of, and in order to do this he develops the judgement-dependent account. However, at the end of this chapter I shall argue that Wright's account fails because it ends up being forced to say that different conflicting judgements (those made at different times) can constitute the same fact, which is deeply problematic.

In contrast to Wright, McDowell argues that there is no need to develop anything like the judgement-dependent account, and that once we have recognised certain common sense facts about meaning and our mental life the question "How is meaning possible?" will no longer seem pressing. In the second of these two chapters I shall turn to evaluating McDowell's account, in the light of the demands upon a satisfactory response to the rule-following problem which the first reveals.

2.1 Kripke's sceptical challenge

In *Wittgenstein on Rules and Private Language*² Kripke (1982) presents an interpretation of Wittgenstein's rule-following argument in *Philosophical Investigations* (Wittgenstein, 1953/2001). At the core of this argument is the fact that we naturally view meanings to

¹Although Wright's book was published before Kripke's, I'm looking at Kripke's first because it is far more straightforward for setting up the basis of the debate, and it's easier to understand Wright's early position in the context of Kripke.

²Kripke notes in *Wittgenstein on Rules and Private Language* (Kripke, 1982, p. ix) that the position he defends here is neither his own, nor meant to perfectly describe Wittgenstein's. Because of this the position developed in this book is often written about as belonging to Kripkenstein or Kripke-Wittgenstein. However, in this chapter I have no need to present Kripke's own views independent of what he writes in *Wittgenstein on Rules and Private Language*, and so shall refer to the narrator of the book simply as Kripke.

be normative; grasp of a word's meaning involves knowing how we *should* use the word, and in what situations it would or wouldn't be appropriate to use. Someone who used the word 'cow' to refer to both cows and horses would be said to have failed to grasp the meaning of the word (assuming there's no other explanation, such as the use being part of a joke). Kripke argues that the normativity of meaning leads to serious problems which undermine realism about meaning.

Kripke's argument for this starts with the example of our understanding of '+'. Imagine you were asked to perform an addition higher than any you'd done before, say '68+57'.³ You would naturally assume that by answering '125' your answer is in faith with your previous understanding of '+'. Kripke's central move in this book is to introduce a sceptic who questions this assumption, and to demonstrate that there is no way to answer the sceptic head on. The sceptic introduces the function *quus*, where *quus* has the same output as *plus* for all numbers smaller than 57, but when given any number greater than 56 as an input the result is always 5. This is problematic because if you've never used '+' with a number bigger than 56 then there will be nothing in your past behaviour, or mental life, involving '+' to indicate which function you meant by '+' — your behaviour would be the same regardless of whether you meant *plus* or *quus* when you used '+'. The sceptic then challenges you to provide justification for the claim that when you used '+' in the past you meant *plus*, and not actually *quus*.

Kripke claims that any successful response to the worries generated by the introduction of the *quus* function must satisfy both of the following two constraints:

The extensional constraint: The response to the sceptic must answer the question of whether there is "any *fact* that I meant plus, not quus, that will answer his sceptical challenge?" (Kripke, 1982, p. 11, emphasis in original). To satisfy this constraint we need to be able to "give an account of what fact is it (about my mental state) that constitutes my meaning plus, not quus" (Kripke, 1982, p. 11). Since the meaning of an expression determines its extension, the fact which constitutes an expression's having a particular meaning must also determine its extension. So, in the '+' example, whatever fact constitutes '+' meaning *plus* and not *quus* must determine '125' and not '5' being the answer to '68+57'.

The normative constraint: Our response to the sceptic must explain if "I have any reason to be so confident that now I should answer '125' rather than '5' "? (Kripke, 1982, p. 11). This constraint demands that any direct response to the sceptical challenge be able to "show how I am *justified* in giving the answer '125' to '68+57' "

³Obviously most people have performed sums involving far larger numbers than 57, but smaller numbers serve for simpler examples. What's key here is that for any speaker there's always going to be some sums which involve numbers larger than those they've ever added — there's always going to be an example of this type, but in an actual example the numbers involved may just have to be quite large.

(Kripke, 1982, p. 11, emphasis added) — it needs to be able to explain how the fact involved in satisfying the extensional constraint *justifies* a speaker giving one answer rather than another. This means that the fact which makes it the case that ‘+’ means *plus* must be accessible to the speaker so that it can explain how they are *justified* in answering ‘125’ to the question ‘What’s $57 + 68$?’ Also, this fact must be distinct from the speaker’s own opinion about how they should use language. We want to make sense of the possibility of mistakes, and so there must be a distinction between usage which merely *seems* right to the speaker, and that which actually *is* right.

It’s important to note that at this stage Kripke’s not developing a sceptical challenge to arithmetic itself. He points out (Kripke, 1982, pp. 12–13) that he’s happy to take for granted that that 68 plus 57 actually is 125 ; he’s not doubting that there is such a thing as the addition function, instead he’s merely questioning how the ‘+’ sign refers to it. The challenge is not to explain addition, but to justify our claim that by our use of the ‘+’ sign that we actually mean the function *plus* rather than *quus*.

Another essential clarificatory point is that the problem is not specific to arithmetical functions, nor even just to functions, but applies to all of language. Kripke (1982, p. 19) gives an alternative example involving two possible meanings of the predicate ‘table’. In this case the sceptic argues that by your uses of ‘table’ in the past, you could have meant either *table* and *tablair* (where a *tablair* is anything which is a table that is not found at the bottom of The Eiffel Tower, or a chair found at the bottom of The Eiffel Tower). Just like the speaker who’s never previously used ‘+’ for numbers greater than 56 , anyone who’s not used the word ‘table’ to refer to an object at the bottom of The Eiffel Tower will have no evidence that when they used the word ‘table’ in the past they didn’t actually mean *tablair*. And like the *quus* problem this example can easily be modified to fit any speaker since no one’s used the word ‘table’ in all possible situations.

Finally, it’s important to recognise that Kripke isn’t merely aiming to make the epistemological claim that we can’t know if we’re using a word in the same way as we were in the past. Instead his strategy is to argue for this epistemological claim, and then use it to justify a metaphysical conclusion. He argues that there is no way to answer the sceptic even if we assume knowledge of all your previous behaviour and entire mental life; there’s nothing in your past behaviour or mental history which demonstrates that by ‘+’ you meant *plus* rather than *quus*. He argues that if this is the case then even an omniscient God would fail to know which meaning was associated with your past use of ‘+’. And if such a God is unable know what you meant by *plus* then we can draw the metaphysical conclusion that “There can be no fact as to what I mean by ‘plus’, or any other word at any time.” (Kripke, 1982, p. 21).

2.1.1 Dispositions

Kripke continues by going through a number of potential responses to the sceptic's challenge and showing why they fail. The most significant of these is the appeal to dispositions (Kripke, 1982, pp. 22–33). We would expect that if I had asked you 'What's $68 + 57$?' in the past then you wouldn't have answered '5', but would actually have been disposed to answer '125'. If this is true, and for all sums you were disposed to give an answer in line with *plus* rather than *quus*, then we can conclude that you must have actually understood '+' to mean *plus*. And in response to the sceptic's challenge to provide a fact which that determines that by '+' you meant *plus* rather than *quus*, we can refer to the fact that you were disposed to *plus*. Kripke argues that this response fails to satisfy both of the two constraints which he claims a successful response to the sceptic must meet.

He argues that this response fails to meet the extensional constraint because our dispositions are finite. There are infinitely many numbers so large that a human being would have no dispositions for how they'd respond when asked their sum because they would die of old age before they were able to perform the calculation. Dispositions would have no power in demonstrating that a speaker didn't associate some deviant *quus*-like meaning with '+' if this *quus*-like function only differed from *plus* for numbers larger than those that could be calculated in a human lifetime.

Dispositions are also unable to satisfy the second constraint and provide *justification* for responding '125' to '68+57'. There are two problems here, both of which involve the fact that speakers can be disposed to make mistakes. The first involves the possibility for speakers to be disposed to make systematic mistakes; if a particular speaker always got confused when adding 8 and 7 that doesn't necessarily mean that they don't mean *plus* by '+'. Dispositionalism can't make sense of this possibility, since all that there is to meaning *plus* by '+' is having the right dispositions. The other problem is that facts about how a speaker will be disposed to use an expression can come apart from how the speaker *should* use the expression. If in a particular instance a speaker was disposed to answer '5' to '68+57' that doesn't necessarily mean that on that occasion by '+' they meant *quus*; they could have actually meant *plus* and just made a mistake. If a speaker means *plus* by '+' then when they are asked 'What's 68+57?' they *should* respond with '125', but it doesn't necessarily follow that that's what they will answer, or what they will be disposed to answer in a particular instance. And so the mere fact that you are disposed to give a particular response when asked to perform an addition doesn't explain how you are able to know what response you *should* give; dispositions can't play the required normative role.

Kripke also explores an attempt to use dispositions which attempts to get round the problem that we sometimes make mistakes by factoring the mistakes out. A crude version of this would be to claim that a speaker means *plus* by '+' if and only if when they do

not make a mistake they will be disposed to *plus*. The problem with this solution is that dispositionalism doesn't provide us with an account of what a mistake is. We would only know to treat a speaker's response of '5' to '68+57' as a mistake if we presume that they are plussing. If they are in fact quussing then that answer would not be a mistake. How can dispositions distinguish answers that are mistakes from the other answers without already presuming which function the speaker has in mind?⁴

Essentially, the problem with dispositions is that they are merely descriptive, whereas meanings are prescriptive. Facts about what you are disposed to say tells us what you would say in any particular circumstance, but they don't tell us what you *should* say. What's key to meaning *plus* by '+' is that you *should* respond '125', the fact that sometimes you'll be disposed to give a different answer doesn't change that.

2.1.2 Internal states

Qualia

Another attempt at a response to the sceptic which Kripke rebuts is the appeal to internal states, such qualia or mental images (Kripke, 1982, pp. 40–41). For example, what if meaning *plus* by '+' has its own special quale which we can know directly by introspection. Introspection of qualia allows us to unproblematically know whether we're having a headache or being tickled; can it also enable us to know whether we are plussing or quussing? Kripke responds to this suggestion by arguing that even if there were a special quale associated with each meaning (which from phenomenological introspection seems highly dubious) it wouldn't help:

If I think that the headache [Kripke's example of the special quale for *plus*] indicates that I ought to say '125', would there be anything about it to refute the sceptic's contention that, on the contrary, it indicates that I should say '5'? (Kripke, 1982, p. 42).

The problem here is that even if a unique quale was experienced when using '+' it wouldn't satisfy the normative constraint because having a particular quale wouldn't tell us how we ought to answer when we come across an addition problem. How would experiencing a certain quale when we saw the '+' symbol tell us that when coming across '68+57' we should answer with '125' and not '5'? Qualia can't tell us what mathematical answers are correct and incorrect, and so they lack the required normativity which is essential for responding to the sceptic's challenge.

⁴As I noted, this is a particularly crude version of the attempt to save the dispositional solution. More sophisticated attempts concentrate on the dispositions of speakers in ideal conditions. For a description of such attempts, an explanation of why these more sophisticated attempts also fail, and a criticism of Kripke's argument against dispositions here, see Boghossian's "The Rule-Following Considerations" (Boghossian, 2002, §§19–25).

Mental images

Kripke next turns to look at possibility of mental images being used to respond to the sceptic (Kripke, 1982, pp. 20, 42–43). Consider the word ‘green’. What if whenever I use the word ‘green’ an image of a green colour patch comes to mind, wouldn’t this show me how to use the word? For example, consider looking at a green field and wondering whether to apply ‘green’ to it. The idea here is that I conjure up the mental image of a green colour patch, compare it to my visual experience of the field, and then check if they’re the same colour. Kripke responds to this view by arguing that the sceptic can suggest an alternative understanding of ‘same colour’ which presents the same problems as *quus* did. Consider the alternative meaning of ‘same colour’, *quame colour*, where two things are the *quame colour* if and only if they are the same colour or one of them is a particular shade of brown (which, like numbers bigger than 56 in the *quus* example, you’d never actually come across before). As with the *quus* example, there would be nothing in your past behaviour or mental life to indicate which meaning you’d associated with ‘same colour’. This means that even if you had always brought a green mental image to mind when you used the word ‘green’ you wouldn’t be able to demonstrate that you meant the same thing by ‘green’ as you do now, since when using the mental image you could have actually been checking whether it was the *quame colour* as the things you encountered in the world.

To get at the core of what’s going wrong let’s follow Kripke in looking at a second example involving shapes. I’m going to explore the point using our understanding of the word ‘triangle’. If we associate the mental image of a triangle with the word ‘triangle’ then we would expect something like the following to tell us whether an object, x , should be called a ‘triangle’:

‘Triangle’ applies to x if and only if x looks like \triangle .

And, as above, the sceptic would respond by suggesting alternative understandings of ‘looks like’. Consider the following:

x quooks like \triangle if and only if
 ((the year is 2013 or earlier and x looks like \triangle), or
 (the year is greater than 2013 and x has one side more than \triangle)).

Once again, there would be nothing in our past behaviour, or mental life, to indicate whether we understand ‘looks like’ to mean *looks like* or *quooks like*.

The problem which both of these examples get at is that we can use mental images differently. Kripke argues that a mental image alone doesn’t immediately tell us how to

use a word, but requires an interpretation in order to do so. An interpretation of a mental image is a rule which tells you how a word associated with that image should be used — it gives you the normative significance of the mental image. (For example, “ ‘Triangle’ applies to x if and only if x looks like \triangle ” is a rule for interpreting the mental image \triangle to know how to use the word ‘triangle’.)

If mental images require interpretation to tell us how to use words this means that simply knowing that the mental image which came to mind when using a word has stayed the same over time doesn’t show that our meaning has stayed the same. We have no way of telling if we have used the mental image with the same interpretation, and so no way of knowing if the meaning hasn’t changed. For example, the fact that we use a mental image of a triangle doesn’t tell us whether the standard of correctness of the word ‘triangle’ is set by checking if the object *looks like* \triangle or *quooks like* \triangle .

Of course, we could attempt to fix the meaning of ‘looks like’ in the same way we fixed the meaning of ‘triangle’, by using a further mental image. Perhaps when understanding ‘looks like’ we bring to mind a mental image that represents lining up all the sides of a shape with that of another, wouldn’t this protect us against the possibility that we may have used ‘looks like’ to mean *quooks like*? But the same problem will recur; this mental image will itself require interpretation to tell us how to use it to give ‘looks like’ a meaning, and the sceptic can suggest a *quus*-like understanding of the interpretation of this further mental image. No mental image alone can have a single determinate interpretation, and any attempt to use a further mental image to fix the interpretation would itself require interpretation, and so we get an infinite regress.

***Sui generis* mental states**

Kripke also briefly considers the idea that there could be a mental state which is completely unlike qualia, mental images, or any other type of mental state, but is *sui generis*:

Perhaps [meaning *plus* by ‘+’] is simply a primitive state, not to be assimilated to sensations or headaches or any ‘qualitative’ states, nor to be assimilated to dispositions, but a state of a unique kind of its own. (Kripke, 1982, p. 51).

He raises two objections to such a view, firstly the problem that it:

seems desperate: it leaves the nature of this postulated primitive state — the primitive state of ‘meaning addition by “plus” ’ — completely mysterious. It is not supposed to be an introspectible state, yet we are supposedly aware with some fair degree of certainty whenever it occurs. For how else can each of us be confident that he *does*, at present, mean addition by ‘plus’? (Kripke, 1982, p. 51, emphasis in original).

Kripke’s second objection to this view is that the state would have to be finite, because it is contained in our finite minds. But if that is so then how could it determine the correct usage of ‘+’ for all numbers?

I will argue in the next chapter that the correct response to Kripke's sceptical challenge is McDowell's. McDowell's response runs roughly along similar lines as the idea that the facts which constitute meaning are *sui generis* mental states. For that reason I'm going to save detailed exploration of this view, and any evaluation of Kripke's criticism of it, until then.

In *Wittgenstein on Rules and Private Language* Kripke (1982) also considers a few other attempts to solve the problem — such as using the behaviour of machines to fix the meaning, considerations of simplicity, and appeals to Platonism — but argues that none of them work and so concludes that there's no way to respond to the sceptic's challenge head on.

2.1.3 Kripke's sceptical solution

If there's no way to respond to the sceptic head on, then at first it may seem that we are left with the following conclusion:

The sceptical argument, then, remains unanswered. There can be no such thing as meaning anything by any word. Each new application we make is a leap in the dark; any present intention could be interpreted so as to accord with anything we may choose to do. So there can be neither accord, nor conflict. (Kripke, 1982, p. 55).

Such a result is clearly disastrous: if we lose the notion of accord and conflict then we lose the normativity of meaning, we can use words however we like and so we would be forced to conclude that all language is meaningless.

Kripke argues that although there's no way to directly respond to the sceptic the prospects for meaning aren't quite so bleak as they may first appear. To clarify the nature of his response he distinguishes between "sceptical" and "straight" solutions to problems of this kind. *Straight solutions* attempt to meet a sceptical argument head on and refute it. In this case that would involve finding a fact which determines what you mean by a word which could demonstrate that you have the same meaning in mind as you did in the past, and would be able to tell you what to do in new instances. In contrast, *sceptical solutions* concede that the sceptic's key conclusion is correct — in this case, the conclusion that we cannot find a fact which constitutes an expression meaning one thing rather than another — but then attempt to go on to show that this doesn't have the disastrous consequences for our practice that it may at first appear to. In the case of meaning, Kripke's sceptical solution involves his conceding to the sceptic that there is no *fact* which makes it the case that you mean one thing or another, but he denies that we must conclude that all language is meaningless.

But if he accepts the sceptic's conclusion that there are no facts which constitute meaning, how does Kripke avoid the conclusion that all language is meaningless? For

language to have meaning there has to be a distinction between correct and incorrect use, this is why it makes sense to talk about language in terms of following rules. And not only do we need a distinction between correct and incorrect usage, but it's also essential to acknowledge the fact that sometimes we make mistakes in our understanding of what use is correct — there is a distinction between the use which *seems* right to us from what actually *is* right. This means that we need to be able to distinguish between 'Jones' use of '+' is correct' and 'It seems to Jones that Jones' use of '+' is correct' — we need to be able to draw the seems right / is right distinction. But if there is no fact which makes it true that Jones' use of '+' is correct then how will we draw this distinction? If there are no facts which constitute meaning then it seems that all we are left with is whatever Jones happens to think about how '+' should be used, and then it will be impossible to draw the seems right / is right distinction.

Kripke's solution is to turn to the assertion conditions of these sentences. He accepts that statements like 'Jones' use of '+' is correct' and 'It seems to Jones that Jones' use of '+' is correct' are never strictly speaking true or false (since there are no facts which can make them true or false), and so we can't distinguish between statements like these using their truth-conditions. But Kripke points out that despite this we will still feel warranted in asserting sentences such as these in different cases, so they will still have differing assertion conditions, and we can use this to draw the essential seems right / is right distinction.

Normally the assertion conditions of a sentence are linked to its truth-conditions, this means that we use knowledge of the truth-conditions of the sentence to reason about what the assertion conditions are — when we would be warranted in asserting them. But in this case the sentences do not have truth-conditions, and so Kripke claims that we won't be able to reason about what the assertion conditions *should* be, instead all we can do is *describe* what their assertion conditions are — describe the cases in which people would say that we would or would not be warranted in asserting them. People would say that the cases in which we are warranted in asserting 'Jones' use of '+' is correct' and 'It seems to Jones that Jones' use of '+' is correct' are different. This means that they have different assertion conditions, and we can draw on these differences in order to draw the distinction between the cases in which we would say 'Jones' use of '+' is correct' and the cases in which we would only be happy to say 'It seems to Jones that Jones' use of '+' is correct'. We can do all this without there being any fact which constitutes Jones meaning *plus* by '+'.

Kripke starts by considering speakers in isolation. In ordinary cases we would simply ignore the worry that someone may be quassing instead of plussing, we do so “without justification” (Kripke, 1982, p. 87), and “blindly” (Kripke, 1982, p. 87). He argues that this isn't wrong, because our language game simply licenses a speaker to give “the answer that strikes him as natural and inevitable” (Kripke, 1982, p. 88). And so in normal cases

it's perfectly fine to assert that 'I am using '+' correctly', or 'I mean the same thing by '+' as I did yesterday'. (Remember, Kripke is aiming to merely *describe* our practice. So in this case the mere fact that our language game licences someone to judge that she has used language correctly is enough. Kripke does not want to involve any underlying fact which makes it the case that someone meant the same thing by '+' as they did yesterday.)

What's problematic about considering speakers in isolation is that it doesn't enable us to draw the seems right / is right distinction. When considering Jones in isolation, the conditions in which Jones would be warranted in asserting 'Jones' use of '+' is correct' and 'It seems to Jones that Jones' use of '+' is correct' would be the same. These sentences could never come apart. All there is to go on is whether Jones' use of '+' *seems* right to him, and there will be no cases in which he'd be prepared to assert that his use seems right to him but not assert that it is right. (If there is no fact which constitutes meaning then when distinguishing between these sentences all we have to go on is their assertion conditions.)

Kripke argues that we can open up the gap between usage seeming right and being right by looking instead at speakers within a linguistic community. He argues that other speakers provide the "justification conditions for attributing correct or incorrect rule following to the subject" (Kripke, 1982, p. 89). If you asked someone 'What's 68+57?' and they responded with '5' then our speech community would judge that they had got the sum wrong. If someone consistently gave answers to addition problems which differed bizarrely from our own then we would judge that they were not following the same rule as us, and that they may well associate a different meaning with '+':

We say of someone else that he follows a certain rule when his responses agree with our own and deny it when they do not (Kripke, 1982, p. 92).

This opens up a gap between the cases in which we'd be warranted in asserting 'Jones' use of '+' is correct' and 'It seems to Jones that Jones' use of '+' is correct'. This means that we have the required seems right / is right distinction, without making use of any fact which constitutes Jones' use of '+' being correct.⁵

⁵It's important to point out that Kripke is not using the assertion conditions of sentences about a speaker's use of language to give us *facts* about whether the speaker's usage is correct. If Jones' use of '+' seems to be in line with the rest of our community then we would feel warranted in asserting that 'Jones uses '+' correctly', But, as we saw above, Kripke is not claiming that we *ought* to make this assertion, he's merely *describing* the fact that we would make this assertion in this case. The sentence 'Jones uses '+' correctly' is not *true* in the case in which his usage is sufficiently similar to our own (since the sceptic has shown that there are no facts in which such a truth could consist); it's simply that we would feel warranted in asserting it. And Kripke claims that we only need the fact that what the community will feel warranted to assert about a speaker's language use can come apart from what seems right to the individual to preserve the essential seems right / is right distinction.

2.2 Responses to Kripke

2.2.1 McDowell

As we've just seen, Kripke attempts to save meaning from disaster whilst accepting the sceptic's conclusion that there is no fact which determines what you mean by a word. He does this by turning to look at speakers in the context of their community, which enables him to save the seems right / is right distinction. But McDowell argues that by accepting the sceptic's conclusion Kripke's position is doomed to lose grasp of meaning altogether:

It is natural to suppose that if one says "There is no fact that could constitute it being the case that P", one precludes oneself from affirming that P; and this supposition, so far from being a distinctively "realist" one, plays a central role in the standard arguments *against* "realism". Given this supposition, the concession that Kripke says Wittgenstein makes to the sceptic becomes a *denial* that I understand the "plus" sign to mean one thing rather than another. And now — generalizing the denial — we do seem to have fallen into an abyss: "the incredible and self-defeating conclusion, that all language is meaningless" [(Kripke, 1982, p. 71)]. It is quite obscure how we could hope to claw ourselves back by manipulating the notion of accredited membership in a linguistic community. (McDowell, 1984/1998o, p. 228, emphasis in original).

What's driving McDowell's argument here is the claim that:

It is natural to suppose that if one says "There is no fact that could constitute it being the case that P", one precludes oneself from affirming that P. (McDowell, 1984/1998o, p. 228).

If this, apparently "natural", supposition were true then it's clear that Kripke's position would be in serious trouble. Kripke accepts the sceptic's claim that there is no fact which determines what you mean by a word, if we accept McDowell's "natural" supposition then this would mean that we could never affirm any claims about what people mean. Kripke relies upon the community's assertions about what people mean to save the seems right / is right distinction, and without them he wouldn't be able to save the notion of normativity; Kripke relies on the fact that we make assertions like 'Jones means addition by '+'', and 'Jones means the same thing by '+' as the rest of our community' to draw the distinction between a usage seeming right and it actually being right. If it would never be correct to make such assertions then Kripke would have nothing left to save the normativity of meaning.

However, there seems to be good reason to doubt the truth the "natural" supposition, or at the very least McDowell needs to provide us with substantial argument in order to justify it, and there is no such argument for it here. McDowell simply asserts it and claims that many critics of realism make use of it. Although this may well be true there are also many non-realist positions that depend upon the rejection of this claim.

For example, consider the various non-cognitivist positions in metaethics, such as Ayer's emotivism in *Language, Truth, and Logic* (Ayer, 1936), and Blackburn's quasi-realism in *Spreading the Word* (Blackburn, 1984b).⁶ For moral claims — such as 'murder is wrong' — non-cognitivists hold that there are no facts which constitute such claims being true because they don't think that such sentences should be evaluated in terms of their truth or falsity. Instead they hold that these moral claims actually express non-cognitive states like desires and emotions. For example, metaethical non-cognitivists hold that when you say that 'murder is wrong', what you are actually doing is expressing your *disapproval* of murder. But, despite the fact that there is no *fact* that constitutes it being the case that murder is wrong, they don't think there is anything wrong with affirming that 'murder is wrong'. And so it's wrong for McDowell to simply claim, without any supporting argument, that you can't affirm P if you don't think there's any fact which could constitute it being the case that P.

Although McDowell offers no argument to support his claim here there are many well known serious problems for non-cognitivism, such as the Frege-Geach problem.⁷ Also, in later papers such as "Values and Secondary Qualities" (McDowell, 1985/1998m) and "Projection and Truth in Ethics" (McDowell, 1987/1998l) McDowell does himself attack non-cognitivist positions. But when I evaluate McDowell's arguments against non-cognitivism in the fifth chapter I eventually conclude that they are not particularly convincing, so I shall not be concerned with them here.

In addition to non-cognitivists, moral error theorists also reject the claim that there are any facts which constitute moral facts being true, but they often hold that we can continue to affirm moral claims. These views endorse a cognitivist understanding of moral talk; they claim that moral judgements express beliefs, but argue that such beliefs are false because there are no moral facts. Although Mackie's insistence (in *Ethics: Inventing Right and Wrong* (Mackie, 1977)) on continuing to make moral assertions after endorsing an error theory was rightly criticised, more recently moral fictionalists (such as Joyce in *The Myth of Morality* (Joyce, 2001)) have more rigorously defended endorsing an error theory and yet continuing to moralise. However, this is a much more recent metaethical development, and it's reasonable that McDowell did not consider this move when he wrote the above passage.

⁶There is in fact some controversy over whether it's fair to describe to Blackburn's position as "non-cognitivist". I shall look at this issue when I look at Blackburn's metaethics in chapter 5.

⁷For the Frege-Geach problem, see Geach's papers "Ascriptivism" (Geach, 1960) & "Assertion" (Geach, 1965), and §5.2.1 of chapter 5.

2.2.2 Wright

Wright's criticism also focuses on Kripke's claim that there is no fact which determines what you mean by a word and his shift to looking at the conditions in which members of a speech community would assert that a speaker had used a word correctly. Wright argues that if Kripke's sceptical argument can be used to show that there is no fact which constitutes associating a particular meaning with a word, or using a word correctly, then an argument along those lines can also be used to show that there are no facts about when your community would assert you'd used a word correctly. If there are no such facts then your community's view of your usage cannot be used to save the normativity of meaning.

Kripke's sceptical argument leads him to conclude that there can be no facts which constitute associating one meaning with a word rather than another. For example, consider the following two sentences:

- I mean the same thing by '+' as I did yesterday.
- The meaning of '+' dictates that when asked 'What's $68+57$?' I should respond with '125'.

If Kripke's sceptic is correct then there are no facts which make it the case that I meant the same thing by '+' yesterday as I do today, or which make it the case that I should give a particular answer to addition problems. So does this mean that such sentences are in fact meaningless? But if sentences about meaning themselves are meaningless then what hope can there be for the rest of language?

As we saw above, Kripke responds to this problem by arguing that we can save the normativity of meaning — the fact that a sentence's use can be correct or incorrect, and that a particular speaker can make mistakes about this — by looking to the community's judgements about whether a speaker's use of a sentence was correct (the assertion conditions of such sentences). And so Kripke concludes that the lack of facts about meaning doesn't prevent us from drawing the seems right / is right distinction.

Wright's argument against this conclusion goes as follows:

what status, once the skeptical argument is accepted, is supposed to be possessed by the sort of account adumbrated by Kripke of the assertion conditions of statements about meaning and understanding. Could it yesterday have been true of a single individual that he associated with the sentence 'Jones means addition by '+' the sort of assertion conditions Kripke sketches? Well, if so, that truth did not consist in any aspect of his finite use of that sentence or its constituents; and, just as before, it would seem that his previous thoughts about that sentence and its use will suffice to constrain within uniqueness the proper interpretation of the assertion conditions he associated with it only if he is granted correct recall of the content of those thoughts — exactly what the sceptical argument does not grant. But would not any truths concerning

assertion conditions previously associated by somebody with a particular sentence have to be constituted by aspects of his erstwhile behaviour and mental life? So the case appears no weaker than in the sceptical argument proper for the conclusion that there *are* no such truths; whence, following the same routine, it speedily follows that there are no truths about the assertion conditions that any of us presently associates with a particular sentence, nor, *a fortiori*, any truths about a communal association. (Wright, 1984, p. 770, emphasis in original).

Wright's concern here focuses on the truth-conditions of sentences about the assertion conditions of sentences about meaning. This is a rather complicated problem, so let's introduce an example:

Yesterday Kirsten held that it is permissible to assert that 'Jones means addition by '+' ' if and only if Jones' use of '+' seems to accord with that of his community.

Wright asks whether it can be true that Kirsten associated such assertion conditions with that sentence. So, what would associating these assertion conditions with that sentence consist in? Well, if we follow the techniques of Kripke's sceptic it must consist in some aspect of Kirsten's past behaviour or mental life. But now what if we introduce a *quus*-like alternative:

Yesterday Kirsten held that it is permissible to assert that 'Jones means addition by '+' ' if and only if

((Jones is not in a philosophy seminar room and Jones' use of '+' seems to accord with that of his community), or

(Jones is in a philosophy seminar room and Jones says he knows what he means)).

If Jones has never stepped foot in a philosophy seminar room then there will be nothing in Kirsten's past behaviour or mental life to demonstrate which assertion conditions she associates with 'Jones means addition by '+' ', and so, following Kripke's sceptical argument, no fact about which assertion conditions she associates with the sentence. This is what justifies Wright's conclusion that:

there are no truths about the assertion conditions that any of us presently associates with a particular sentence, nor, *a fortiori*, any truths about a communal association (Wright, 1984, p. 770).

Kripke's sceptical solution depends upon there being facts about what assertion conditions speakers, and communities, associate with sentences about meaning. He needs these facts in order to use these assertion conditions to draw the seems right / is right distinction. If there are no such facts then he is unable to ground the normative aspect of meaning in them, and Kripke's solution is doomed to fail.

2.2.3 Summary

Kripke's sceptic challenged us to find a fact which could constitute having a meaning in mind. Such a fact needs to satisfy two constraints: it must be able to determine the correct extension of our words — the fact which makes it the case that I use '+' to mean *plus* must determine how to correctly use *plus* in all cases —, and it must be normative — we need to be able to explain how I am able to know that I *should* use '+' in the correct way in each case. Kripke claims that there are two ways to respond to a sceptical challenge of this sort: either with a "straight" solution which finds a fact that meets these constraints, or a "sceptical" solution which acknowledges that there is no such fact, but finds a way to limit the damage so that the very notion of meaning isn't undermined.

Kripke thinks there is no "straight" response to the sceptic, and so develops a "sceptical" solution. Despite thinking that there is no fact which constitutes having a meaning in mind, he still needs to be able to make sense of the normativity of meaning; he thinks it is essential to the very idea of meaning that there are correct and incorrect ways of using language. In addition, we need to be able to make sense of the distinction between use which merely seems correct, and that which is actually correct, since speakers can sometimes make mistakes in their judgements of how they should use a word. Kripke attempts to account for all of this by making use of the community's judgements about a speaker's use of language. He points out that the community would judge a speaker to be using a word correctly if they use it in a way which aligns with their own use. There may be no facts about how we should use language, but this enables us to continue to make sense of our judgements about a speaker's use of language as being right or wrong. However, as we have just seen, Wright shows that this solution fails because it depends upon there being facts about when the members of the community would make judgements about a speaker's usage, but the original sceptical challenge can easily be reapplied to also undermine these facts.

As things stand this is disastrous. We have no account of any fact which could constitute having a particular meaning in mind, but without such a fact there seems to be no way to account for the idea that there can be right and wrong ways to use language. We seem left with "the incredible and self-defeating conclusion, that all language is meaningless" (Kripke, 1982, p. 71).

2.3 Wright's communitarian solution

Next I'm going to turn to look at two responses to Wittgenstein's rule-following problem developed by Crispin Wright, both of which play significant roles in McDowell's own work. I'm going to start by looking at Wright's early view, as laid out in *Wittgenstein on the Foundations of Mathematics* (Wright, 1980) where he put forward a reading of

Wittgenstein which, like Kripke's, responded to the rule-following problem by turning to the community. Wright soon dropped the view advocated in this book and instead responded to Kripke's sceptical challenge by arguing that Kripke's is guilty of making a reductionist assumption. Wright claims that if we drop that assumption then we can give a successful account of meaning by drawing a parallel with intentions and by claiming that the facts about meaning are response-dependent. The earliest version of that view which I'll look at is from his 1984 paper "Kripke's Account of the Argument Against Private Language" (Wright, 1984), and I'll trace the development of the view up to the paper he wrote in 2000, "Self-Knowledge: The Wittgensteinian Legacy" (Wright, 1998/2000).⁸ In this section I shall evaluate Wright's position developed in *Wittgenstein on the Foundations of Mathematics*, and shall conclude by agreeing with Boghossian's argument that it fails because it can't account for the possibility of the community going wrong.

As with Kripke, an essential aspect of Wright's view of meaning in *Wittgenstein on the Foundations of Mathematics* (Wright, 1980) is that meaning is normative; there are ways we should and should not use words in order to accord with their meaning, and we need to be able to make sense of the distinction between a use of a word *seeming right* to a speaker and it actually *being right*. Answering '5' to 'What's 68+57?' would be wrong because it does not accord with the meaning of the words in the question, only '125' would be the correct response, regardless of whether it *seems* to the speaker that that's what they should say. Similarly any speaker who referred to a horse as a 'cow' would be failing to comply with requirements of the meaning of the word 'cow'.

Wright argues that the normative aspect of meaning becomes problematic when combined with another natural assumption about meaning, which is that understanding a word's meaning involves "grasp of a pattern of application, conformity to which requires certain determinate verdicts in so far unconsidered cases" (Wright, 1980, p. 216). According to this idea "[the truth of statements] is settled, autonomously, and without the need for human interference, by their meanings and the character of the relevant facts" (Wright, 1982, p. 250).

This idea is equivalent to the notion of "objectivity of meaning" which Wright explains when distinguishing between three types of objectivity in the introduction to his book

⁸Although I will only be looking two positions Wright has held on the rule-following problem he has in fact recently started to consider a third response. In "Rule-Following Without Reasons: Wittgenstein's Quietism and the Constitutive Question" (Wright, 2007) he argues that his previous view couldn't give an account of the most fundamental cases of rule-following, and so fails to solve Wittgenstein's problem. He goes on to look at a quietist solution, much closer to McDowell's, although it isn't a position he's particularly satisfied by.

Realism, Meaning, and Truth (Wright, 1993a). There he defines the view as follows:

the meaning of a statement is a real constraint to which we are bound, as it were, by contract, and to which verdicts about its truth-value may objectively conform, or fail to conform, quite independently of our considered opinion of the matter (Wright, 1993a, p. 5).

Wright argues in the introduction to *Realism, Meaning, and Truth* that objectivity of meaning is significant, not merely as a thesis about meaning, but also because of its consequences for our view of truth. Wright (1993a, pp. 5–6) argues that if we reject objectivity of meaning then we must also reject objectivity of truth, which is the view that our sentences can have potentially evidence-transcendent truth-conditions.⁹ According to Wright objectivity of meaning is required by objectivity of truth because without objectivity of meaning there would be no way for the meanings of our statements to “[reach] into regions where we cannot follow” (Wright, 1993a, p. 5) — our statements couldn’t refer to facts which were beyond our ability to verify and so they couldn’t have evidence-transcendent truth-conditions.

In *Wittgenstein on the Foundations of Mathematics* Wright argues that we should reject objectivity of meaning and think that “there is in our understanding of a concept no rigid, advance determination of what is to count as its correct application” (Wright, 1980, p. 21). This means the loss of the idea that there are any facts about meanings — such as how to continue a numerical series, or what we should call a ‘cow’ — independently of what we actually go on to say about the matter. He thinks that we should instead associate correct understanding of a word with keeping in step with members of our speech-community. So if we reject the requirement that meanings be objective then we can retain a place for the seems right / is right distinction by embracing a form of communitarianism. And since objectivity of meaning is required for objectivity of truth this means that if his argument in this book is correct then we will also be compelled to drop objectivity of truth.

2.3.1 Wright’s argument

Wright argues against objectivity of meaning using a reductio argument, looking at how we are forced to view meaning if we assume it is objective, claiming that the resulting picture is unacceptable, and so concluding that we should reject objectivity of meaning. He starts by looking at how we learn the meaning of a word. If we assume that understanding a meaning consists in grasp of an objective pattern (a pattern which determines how you should use the word in all cases, including those not yet encountered by your community)

⁹Objectivity of truth is the third of the three realist commitments which we saw in the previous chapter — the one which Wright argued that the manifestation argument compels us to drop.

then “no explanation of the use of an expression is proof against misunderstanding” (Wright, 1980, p. 216) — no matter how many details we are given when taught a meaning there is always scope to understand it in various ways, all linguistic training is essentially indeterminate. For example, you may have been taught to use the word ‘cow’ only in cases where you’re interacting with cows in fields, it would be compatible with that training to think that cows which are standing on motorways were actually called something else; your training doesn’t determine whether you should understand ‘cow’ to mean *cow* or *cow-not-on-motorway*. This point is essentially Kripke’s sceptic’s *quus* point, only applied to the case of learning meanings, no matter how someone had taught you to use ‘+’ it would always be possible to interpret it have a *quus*-like meaning. Wright argues that if grasp of a meaning is essentially grasp of an objective pattern then when understanding an explanation of a word’s meaning it would be necessary for you to make “a leap, an inspired guess at the pattern of application which the instructor is trying to get across” (Wright, 1980, p. 216) in order to associate a particular objective pattern with the meaning.

If we accept this then that would commit us to an idiolectic picture of understanding — understanding is essentially private to individuals, and grasping the understanding of someone else requires a leap, or a guess. Wright argues that this idiolectical picture of understanding cannot be correct since it doesn’t give us the resources to distinguish between correct and incorrect usage, and prevents us from drawing the essential seems right / is right distinction. If grasping the meaning of a word is essentially private then there is no way for us to distinguish between actually grasping a word’s meaning, and instead grasping a *quus*-like deviant meaning (such as *cow-not-on-motorway*). If meaning is private in this way then there would be no standards for us to check our meaning against other than what seems right to us at the time, and our current belief about a word’s meaning is always going to seem right to us at the time — that’s no standard of correctness at all!

He next turns to communal understanding, can we save objectivity of meaning by associating correct usage with acting in line with your community? According to this account the reason it’s wrong to answer ‘5’ to ‘What’s 68+57?’, and why it’s correct to call cows on motorways ‘cows’ is because that’s how your community acts. But he argues that the same problem simply recurs, just at the level of the community. If what we grasp when we grasp a meaning is understood as an objective pattern with a determinate meaning stretching out to infinity then our community’s usage so far will be hopelessly indeterminate between infinitely many different possible meanings. For all the numbers greater than those our community has added, and for all the places we have never spoken about cows visiting, there will be another *quus*-like objective meaning that our community might mean when they use ‘+’, and so there will be no fact about which objective meaning our community grasps.

2.3.2 Wright's solution

Wright responds to this problem by arguing that we can save the normativity of meaning by dropping the objectivity of meaning. This means rejecting the claim that the facts about meaning are independent of human opinion and reach beyond our dispositions to use language.

If we do this then it is possible to use the community's dispositions to restore the normativity of meaning, at least for individual speakers:

it is a community of assent which supplies the essential background against which alone it makes sense to think of individuals' responses as correct or incorrect ... None of us can unilaterally make sense of correct employment of language save by reference to the authority of communal assent on the matter; and for the community itself there is no authority, so no standard to meet. (Wright, 1980, pp. 219–220).

For the individual we can save the notion of error, and the distinction between correct and incorrect usage of a word, by comparing the individual's use to the dispositions of her community — error occurs when an individual goes out of step with what her fellows would say. It is wrong to answer 'What's $68+57$?' with '5' because our community would typically be disposed to respond with '125'. Similarly, it's correct to call cows standing in all sorts of places, and not only in fields, 'cows', even if so far we've only used the word in a field because that is what we would be disposed to do. Wright claims that this move enables us to draw a distinction between the use of a word being right and merely seeming to be right, and so save the normativity of meaning.

It's important to be clear on how Wright's reading of Wittgenstein differs from Kripke's. Kripke's solution depended upon rejecting the idea that there were any *facts* involving meaning, or how we should use language. Instead, he makes use of the assertion conditions for sentences about meaning — the conditions in which a speaker's community would assert that they had used a word correctly or incorrectly — and uses that to give us the seems right / is right distinction. But, as we saw in §2.2.2, Wright (1984) argues that such a solution can't work because if the sceptical argument can be used to show that there are no facts about meaning then it can also be used to show that there are no facts about when the community would assert an individual to have used a word correctly.¹⁰

Although in his later work Wright advocates a judgement-dependent account of meaning, in *Wittgenstein on the Foundations of Mathematics* he makes no such claim. Instead, he thinks that our understanding of sentences about meaning (such as 'I mean the same

¹⁰We should be careful when considering Wright's argument against Kripke in the context of discussing Wright's position in *Wittgenstein on the Foundations of Mathematics* (Wright, 1980) because that book came out two years before Kripke's *Wittgenstein on Rules and Private Language* (Kripke, 1982), and four years before Wright's attack of it in "Kripke's Account of the Argument Against Private Language" (Wright, 1984).

thing by ‘+’ as I did yesterday’, or ‘The meaning of ‘+’ dictates that when asked ‘What’s 68+57?’ I should respond with ‘125’’) consists in grasp of their truth-conditions. By rejecting the objectivity of meaning all he’s doing is claiming that the truth-conditions of such sentences are determined by how the community as a whole is typically disposed to act. All he’s rejecting is the idea that the facts about meaning can go beyond the community’s dispositions, but, unlike Kripke, he attempts to continue to hold on to the idea that there are such facts.¹¹

2.4 Criticisms of Wright’s communitarian view

2.4.1 McDowell

In “Wittgenstein on Following a Rule” McDowell (1984/1998o, p. 235) argues that Wright’s early response to the rule-following problem fails because the rejection of the objectivity of meaning — the rejection of the idea that the facts about meaning are independent of our community’s views — is actually incompatible with meaning’s normativity.

Wright attempts to preserve the seems right / is right distinction, and save the normativity of meaning, by claiming that we are able to draw the distinction between an individual’s correct and incorrect usages of a word by referring to our community’s dispositions to use the word — our usage is correct when it corresponds to that of our community. But, McDowell argues that this solution falls apart once we turn to look at the community itself:

The picture Wright offers is, at the basic level, a picture of human beings vocalizing in certain ways in response to objects ... There are presumably correspondences in the propensity of fellow members of a linguistic community to vocalize, and to feel comfortable in doing so, that are unsurprising in the light of their belonging to a single species ... But at the basic level there is no question of shared commitments — of the behaviour, and the associated aspects of the streams of consciousness, being subject to the authority of anything outside themselves. (“For the community itself there is no authority, so no standard to meet.” (Wright, 1980, p. 220)) ...

... at [the basic level] Wright’s picture has no room for norms, and hence — given the normativeness of meaning — no room for meaning. ... It is problematic, however, whether the picture of the basic level, once entertained as such, can be prevented from purporting to contain *the real truth* about linguistic behaviour. In that case its freedom from norms will preclude our attributing any genuine substance to the etiolated normativeness Wright hopes to preserve. (McDowell, 1984/1998o, p. 235, emphasis in original).

¹¹It’s important to note that it’s somewhat controversial whether at the time Wright would have actually endorsed this reading of *Wittgenstein on the Foundations of Mathematics* (Wright, 1980). But in the context of comparing this book to Kripke’s *Wittgenstein on Rules and Private Language* (Kripke, 1982) it’s hard not to interpret Wright as a factualist about meaning.

By the “basic level” McDowell seems to mean the linguistic practice of the community seen as a whole (presumably this level is “basic” because it is what grounds the facts about how we should use words). At this level Wright drops any claims to normativity, that only applies when looking at an individual speaker in the context of their community. According to McDowell this is problematic because it is the “basic level” where the real truth about language is to be found, and so if there are no norms at that level then there isn’t really a normative practice at all. This leads him to conclude that “the possibility of going out of step with our fellows gives us the *illusion* of being subject to norms, and consequently the *illusion* of entertaining and expressing meanings” (McDowell, 1984/1998o, p. 235, emphasis in original).

This argument depends upon the assumption that if a community as a whole isn’t answerable to anything independent of itself then the individual members of the community can’t be answerable to any norms themselves — if there are no norms at the “basic level” then there can’t be any norms. But what’s wrong with the idea of members of a community simply being answerable to each other, without the community as a whole being answerable to anything independent of it? Why does there they need to be anything external for the individual’s practice to be normative?

Consider the case of three triplets, James, Joe, and John who agree to always wear the same clothes. Every day they take it in turns to get dressed first, and then the other two follow suit. But then one day John and James get into a bitter argument and so James decides to rebel. The next morning John puts on green socks, Joe follows suit, but James puts on red socks instead. Clearly in this case James, Joe, and John are subject to norms, and that by wearing red socks James has violated these norms. As in Wright’s view of language, they are not “subject to the authority of anything outside themselves”, and at the “basic level” — the level of the group as a whole — there are no norms; whoever gets dressed first can wear any colour of socks they like. But that in no way conflicts with the fact that each individual’s practice is normative, and that certain standards of correctness must be met in order for their behaviour to continue to meet the demands of the norms.¹²

And why shouldn’t the same apply to language? Even if the community as a whole is not subject to any external norms — there’s nothing independent of the community to say whether we should *plus* or *quus* when we come across numbers we’ve never added before. That doesn’t change the fact that when an individual is responding to ‘What’s 68+57?’ they are still subject to norms.

¹²Despite it’s *prima facie* plausibility, there is in fact a certain degree of controversy over whether an argument along these lines actually works, for example see Blackburn’s paper “The Individual Strikes Back” (Blackburn, 1984a).

2.4.2 Boghossian

Boghossian (1989, pp. 354–356) evaluates Wright’s position by testing it against the two constraints Kripke claims a successful response to the sceptical challenge must satisfy. He concludes that it can easily satisfy the normative constraint, but it fails to meet the demands of the extensional constraint. I shall start by looking at his evaluation in terms of the normative constraint.

The normative constraint demands that any response to the sceptic’s challenge must explain how the facts in which meaning consists can provide an individual speaker with reasons to use a word in a particular way — what justifies you giving the answer ‘125’ to the question ‘What’s $68+57$?’ and how do you know that’s the correct response? According to Wright, a word’s meaning is dependent upon the dispositions of the members of your speech community and so in order to know the meaning of a word you’d simply need to know how your community would behave. Boghossian acknowledges that “communal agreement on judgement does usually provide one with some sort of reason for embracing the judgement” (Boghossian, 1989, p. 353). That seems reasonable, we often have to track the behaviour or dispositions of other members of our community, and not just in cases involving meanings. Politicians tend to attempt to develop policies which will be liked by the electorate, bakers aim to make cakes which people will enjoy eating — our ability to recognise and predict the behaviour of the rest of our community seems to be an essential skill, and there’s no reason to suspect that we’d be any worse at doing this in the case of linguistic behaviour.

In order to satisfy the extensional constraint Wright needs to give an account of a fact that constitutes us meaning one thing by a word rather than another. This fact needs to be able to make it the case that I mean, say, *plus* and not *quus* by ‘+’, and must be able to determine the correct extension. At first sight it may seem that Wright would be able to meet this constraint as easily as he does the other, since he holds that the facts of what a word means are simply determined by the community’s dispositions to use the word.¹³ However, Boghossian argues that Wright’s account falls down here because the community’s dispositions will identify the wrong extensions.

Consider the word ‘horse’, in most situations our community will be disposed to apply ‘horse’ only to horses, the problems come in when we start talking about slightly less common situations. For example, consider coming across a horsey-looking cow on a dark night. In that situation I may well be disposed to call the animal a ‘horse’, despite it actually being a cow. We would want to say that by calling the cow a ‘horse’ I’d made

¹³Unlike for Kripke, the fact that these dispositions are finite is not a problem for Wright, since in rejecting the objectivity of meaning he’s rejected the claim that the meaning of *plus* stretches out to numbers so large we can’t even have dispositions regarding how to add them. He claims that the meaning of *plus* is just as limited as our dispositions.

a mistake: ‘horse’ should only be used to refer to horses, not to cows in bad lighting conditions that just happen to look a little horsey. But Wright’s account has difficulty in classifying such judgements as mistakes.

In cases like this it may well be the case that the community as a whole will have the disposition to call such a cow, when encountered on a dark night, a ‘horse’. This disposition may well be general and systematic, because on dark nights the cow looks so much like a horse. But our general disposition to make this judgement doesn’t change the fact that calling such cows ‘horses’ is a mistake. This is problematic for Wright’s account of meaning in *Wittgenstein on the Foundations of Mathematics* because there he equates correct usage with the community’s dispositions, and so he doesn’t have space for the idea that community as a whole could share a disposition to make a mistake. As Boghossian argues:

The communitarian, however, cannot call them *mistakes*, for they are the community’s dispositions. He must insist, then, firm conviction to the contrary notwithstanding, that ‘horse’ means not *horse*, but, rather, *horse or cow*. (Boghossian, 1989, p. 356, emphasis in original).

Since the community as a whole will be generally disposed to refer to any horsey looking cows encountered on dark nights as a ‘horse’ then that means that Wright is committed to holding that ‘horse’ actually *should* be used to refer to such cows, and that that is part of the meaning of the word. This problem is in no way specific to horses and cows, but applies generally to any case where we the community as a whole will have dispositions to make systematic mistakes. And because many of our words will have a number of difficult cases in which we’ll all be disposed to make mistakes, Boghossian concludes that, according to Wright’s view, “none of our predicates have the extensions we take them to have, but mean something wildly disjunctive instead” (Boghossian, 1989, p. 356).

The key problem here is, as McDowell points out in “Wittgenstein on Following a Rule” (McDowell, 1984/1998o), that Wright’s account simply doesn’t have space for the idea that an entire speech community can go wrong. McDowell points out that in *Philosophical Investigations* Wittgenstein (1953/2001) emphasises the fact that there’s a distinction between saying ‘This is yellow’ and ‘This would be called ‘yellow’ by (most) speakers of English’ (McDowell, 1984/1998o, p. 234). Wright can’t make sense of this distinction since, as Boghossian shows, he doesn’t have room for the idea that an entire speech community could go wrong. Wright effectively reduces correctness to communal consensus, but that doesn’t actually fit with the fact that we generally hold that truths about meanings can go beyond mere consensus.

This reveals a new requirement upon responses to the rule-following problem. Not only must we be able to makes sense of the seems right / is right distinction for individuals, we must also be able to do so at the level of entire communities. It’s hard to deny the fact that it’s possible for entire communities to make mistakes in their use of language

(otherwise, as Boghossian showed, we will be forced to accept that the meaning of our predicates are in fact significantly different from what we had thought, and are instead wide disjunctions), and so it seems that any successful account of meaning must be able to account for the possibility of a use of language seeming right to a community, but not actually being right.

So, although Wright has identified a fact as being the one that determines the extensions of words — the fact of how our community will be disposed to use the word — his solution doesn't work because it can't draw the seems right / is right distinction at the level of entire communities. His view seems to get the extensions wrong in any case where the community will generally be disposed to make mistakes. If Wright bites the bullet and simply asserts that in those cases the apparent mistakes are actually how we *should* use the words then his view will be guilty of conflicting with the nigh-on inescapable intuition that whole communities can make mistakes, and that correct usage can be distinct from the community's consensus.

One natural way to respond to this criticism would be to find a way to ignore our dispositions on dark nights, where we sometimes mistake cows for horses, and instead focus only on the dispositions we have under ideal conditions. Wright's judgement-dependent account of meaning works along precisely these lines, and it is to this that we shall turn to next.

2.5 Wright's judgement-dependent account

My evaluation of Wright's response-dependent account of meaning is going to focus on the papers starting with "Kripke's Account of the Argument Against Private Language" (Wright, 1984), and following the progress of his position up to "Self-Knowledge: The Wittgensteinian Legacy" (Wright, 1998/2000). In these papers he argues that Kripke's sceptical argument depends upon a reductionist assumption which Kripke has failed to justify. Once we remove that assumption we open the door to a "straight" solution to the sceptical problem, one which can deal with the fact that we can be disposed to make mistakes by focusing only on the judgements we would make in good conditions. This enables him to draw the essential distinction between a use seeming right and being right, and, unlike his earlier position, he does so without depending upon the speaker's wider community.

2.5.1 Wright's attack of Kripke's inferential assumption

As we saw at the start of this chapter Kripke argues that even with complete knowledge of your previous behaviour and entire mental life we can find no fact which determines what you mean by a word, because there will always be many different potential interpretations

of what you mean that are equally compatible with your past behaviour and mental life.

In “Kripke’s Account of the Argument Against Private Language” Wright (1984, pp. 773–777) argues that in order for Kripke’s sceptical argument to work it needs to block us from responding to the sceptic’s challenge by referring to our previous knowledge of what we meant by the word. At first sight this seems to be a pretty reasonable restriction, since Kripke’s sceptic is directly questioning whether we have such knowledge it would seem strange to respond to the sceptic by referring to that knowledge itself. But Wright argues that knowledge of our meaning could be non-inferential, and if it were then we even if we had knowledge of our meaning then we still wouldn’t be able to respond to the sceptical challenge. He goes on to argue that Kripke needs to supplement the sceptic’s challenge with an argument to show that the knowledge of our meaning cannot be non-inferential, and that he fails to do so.

Wright supports his claim about the significance of the possibility that we could have non-inferential knowledge of our meanings by exploring a parallel sceptical challenge about our former perceptions. Consider a Kripkean sceptic about former perceptions who challenges me to provide an account of a fact which makes it the case that *yesterday, I saw it raining*. As in the case of meaning, in responding to this sceptic I would be allowed to assume complete knowledge about all my past (including my past behaviour and mental life), only I could not presuppose that I had knowledge of what I formerly perceived. This would mean that when attempting to answer the sceptic I couldn’t make use of my memories of what I perceived, but instead would be “restricted to my present *seeming-memories*, the presently available testimony of others, presently available putative traces, like damp ground, etc. and meteorological office and newspaper records” (Wright, 1984, p. 774). The problem is that these facts alone would not be enough to provide a satisfactory response to the sceptic because the obtaining of all of these would be consistent with the possibility that it’s false that *yesterday I saw it raining* — as in the *quus* case there’s always going to be an another possibility which is equally compatible with this limited evidence. This means that the sceptic seems able to conclude that there is no fact about what I perceived in the past, and since this argument would still apply in the future about now, then there is no fact about what I am currently perceiving.

Clearly something has gone wrong here, and Wright argues that what’s wrong is:

the assumption that knowledge of a former perception has to be *inferential*, that the ultimate grounds for such knowledge must reside in knowledge of a different sort (Wright, 1984, p. 774, emphasis in original).

Knowledge of the fact that we had a particular former perception is not inferred from anything beyond the fact that we actually had that perception, and so without access to memories of our former perceptions there is no way to justify the claim that we did have any particular past perceptual knowledge. Wright argues that knowledge of our meaning

could be just like this, or at least that unless Kripke provides an argument to show that knowledge of meaning cannot be non-inferential then we should not be worried by our failure to respond to Kripke's sceptical challenge.

In order to justify the idea that knowledge of our meaning could be non-inferential Wright looks to our common sense understanding of intention. He points out that:

The ordinary notion of intention has it that it is a characteristic of mind . . . that a subject has, in general, authoritative, and noninferential access to the content of his own intentions, and that this content may be open-ended and general, may be related to all situations of a certain kind. (Wright, 1984, p. 776).

Additionally, he points out that:

we can in general make no ready sense of the question, "How did you know?" directed at an avowal of intention (Wright, 1984, pp. 775–776).

which suggests that knowledge of our intentions is non-inferential, and so an inappropriate target of Kripke's sceptical challenge. He claims that meanings are relevantly similar to intentions in this respect, and so seem to be equally suitable to be known non-inferentially.

Wright is initially cautious about fully embracing this response to Kripke, since it could be that the intuitive notion of intention which it depends upon turns out to be radically incoherent. His main point here is simply to show that Kripke's sceptical argument, as it stands in *Wittgenstein on Rules and Private-Language* (Kripke, 1982), needs to be supplemented with an additional argument to show that knowledge of meaning is non-inferential if it is to have any force.

2.5.2 Judgement-dependence

But if knowledge of our meaning is not inferential, then how can it be independent of our dispositions to make judgements about meanings? (As we saw in §2.4.2, Boghossian shows that meaning must be independent of our dispositions to use language, otherwise we won't be able to make sense of the fact that we can be disposed to make mistakes.) Wright's solution is to develop a judgement-dependent account of meaning.¹⁴

The classic example of judgement-dependent concepts are those of the secondary qualities, such as colour. According to judgement-dependent accounts, facts about colours are not true utterly independently of our own affective phenomenology, but neither is it the case that we should endorse an error theory about colour and hold that strictly speaking

¹⁴It's important to note that Wright emphasises (Wright, 1989b, p. 246) that he is intentionally disagreeing with Wittgenstein's own response to the rule-following problem. As we'll see later in this chapter, what Wright regards as Wittgenstein's "official" solution has no need for all the machinery that Wright introduces here, because Wittgenstein endorses a Quietism which Wright claims to be philosophically unsatisfying.

all positive judgements involving colour are false. The judgement-dependent account is able to make both of these claims by holding that the facts about colour are constituted by what we would judge to be true in ideal conditions. For example, it is true that *this rose is red* because in ideal conditions (in normal sunlight, when the observer is not drunk, etc.) we would judge that *the rose is red*. In contrast, facts about shape are commonly held to be judgement-independent because they are utterly independent of how we happen to experience the world, and instead sentences about shapes are made true by the world alone. Wright argues that we can find out whether a concept is judgement-dependent or judgement-independent by testing if it passes the order-of-determination test.

The order-of-determination test

The order-of-determination test is used to know how we should best read the following biconditional equation:

x is Q if and only if (if conditions C obtain then a suitable subject would judge that x is Q).¹⁵

There are two ways of reading this biconditional:

The detective reading: The left-hand side of the biconditional has priority, x 's being Q is an independent, determinate state of affairs which the subject is able to correctly apprehend when conditions C obtain.

The constitutive reading: The right-hand side of the biconditional has priority, x 's being Q is constituted by the fact that that is what suitable subjects will judge when in conditions C .

If the detective reading of the biconditional is appropriate then Q is a judgement-independent property because the facts about x 's Q -ness are made true independently of what subjects will judge. If the constitutive reading is appropriate then Q is a judgement-dependent property, because the facts about x 's Q -ness are constituted by suitable subjects being disposed to make such judgements in the appropriate conditions.

¹⁵Wright points out in "Moral Values, Projection and Secondary Qualities" (Wright, 1988, p. 14, fn. 26) that this formulation of the equation is flawed and doesn't actually work in all cases. For example, there are some plants which change colour when exposed to sunlight; such a plant may currently be a sickly white colour but would be disposed to change to a vibrant green when placed in sunlight. If conditions C involved the plant being exposed to sunlight this would mean that even when the plant is currently white and in the dark it would still be true that "if conditions C obtain then a suitable subject would judge that the plant is green", but clearly we shouldn't conclude that the plant is always green!

We can solve this problem by replacing the above equation with the provisional biconditional:

If conditions C obtain, then (x is Q if and only if a suitable subject judges that x is Q).

But for the purposes of clarity at this stage I shall ignore this complication, and present the equation as a biconditional, as Wright himself often does.

So in the case of colour objects it is true that *the rose is red* because that is what a suitable subject will judge in suitable conditions. In contrast, in the case of shape, it is true that *the table is square* because of the properties of the table itself. We can make sense of facts about an object's shape without needing to make any reference to the dispositions of subjects to make certain judgements about it.

But, for any particular concept, how do we know which reading is appropriate? Wright argues that we should endorse the constitutive reading, and conclude that a concept is judgement-dependent if and only if the following four tests can be satisfied:

(i) that we can construct *a priori* true provisional equations for such judgements; (ii) that the C-conditions of these equations can be substantially specified, in a manner free of the triviality associated with whatever-it-takes formulations; (iii) that the satisfaction of the C-conditions is, in any particular case, logically independent of the details of the extensions of colour concepts; (iv) that no other account is available of what else might determine the extension of the truth-predicate among judgements of colour, of which the satisfaction by the relevant provisional equations of conditions (i)–(iii) would be a consequence. (Wright, 1989b, p. 248).

Wright justifies the four tests as follows:

- i. **The *a prioricity* condition:** The truth of the biconditional must be knowable *a priori* because it should be a fact about the concept itself that it is judgement-dependent.
- ii. **The non-triviality condition:** We need the non-triviality condition because without it all predicates would come out as judgement-dependent. As Wright argues, “it is an *a priori* truth of *any* kind of judgement whatever that, if I operate under conditions which have everything it takes to ensure the correctness of my opinion, then it will be the case that P if and only if I take it to be so” (Wright, 1989b, p. 247, emphasis in original). We can avoid the triviality involved in such “whatever it takes” C-conditions by ensuring that they are “plausible, substantial, [and] non-trivial” (Wright, 1989b, p. 247).
- iii. **The independence condition:** The C-conditions cannot be logically dependent upon truths concerning the property in question because then it wouldn't make sense to say that best opinions *constitute* the fact that the property obtains, since whether or not an opinion was best would itself depend on whether that property obtains.
- iv. **The extremal condition:** There needs to be no other ways of explaining why it's *a priori* that in the C-conditions our judgements co-vary with the facts about that property; the judgement-dependence of the property must be our best explanation

of this co-variance. (Wright justifies this using the example of pain, for which he thinks the other three conditions are satisfied because we are infallible about pain, not because it is judgement-dependent, see “Truth and Objectivity” (Wright, 1992, pp. 123–124).)

In the case of making a colour judgement, such as a rose being red, Wright suggests that the appropriate C-conditions would be something along the lines of:

the surface must be in full view, and in good light, relatively stationary, and not too far away; and the subject must know which object is in question, must observe it attentively, must be possessed of normal visual equipment and be otherwise cognitively lucid, and must be competent with the concept *blue*. In addition the subject must be free of doubt about the satisfaction of any of these conditions — for doubt may lead to an unwillingness to make any judgement (Wright, 1989b, p. 247).

He argues (Wright, 1989b, pp. 247–248) that for colour judgements all of the four tests can be satisfied, and so we should conclude that colour is a judgement-dependent concept.

In the case of shape Wright argues that these tests can’t be satisfied because if the shape is in three-dimensional space then the judgement will need to be made on the basis of several observations from different angles (this is to avoid the appearance of the object’s shape being distorted because we are looking at it from an angle). But for the subject to be able to use multiple observations to judge a shape then the C-conditions must involve the requirement that *x*’s shape doesn’t change between the observations, which is problematic:

we need to ensure that no *change* in *x*’s shape takes place though the period of these several observations. But that calls for some ingredient in the C-conditions of which it is an *a priori* consequence that whatever it is true to say of *x*’s shape at any time during the subject’s observations is also true at any other time within the relevant period. Some independent determinant is therefore called for of what it *is* true to say about *x*’s shape during that period — independent, that is, of the opinion formed by the subject. There is accordingly, it seems, no immediate prospect of a provisional equation for ‘*x* is pear-shaped’ meeting both conditions (i) and (iii) above. (Wright, 1989b, p. 248, emphasis in original).¹⁶

2.5.3 The judgement-dependent account of intentions

Wright next moves on to develop a judgement-dependent account of self-ascriptions of intentions, which he uses to provide the archetype for a judgement-dependent account of

¹⁶Miller reveals serious problems with this argument, see “Primary qualities, secondary qualities and the truth about intention” (Miller, 2009), and “Another objection to Wright’s treatment of intention” (Miller, 2007). However, in this chapter my focus will be on the flaws elsewhere in Wright’s position, so I shall not explore the problems raised by Miller.

meaning.

According to Wright one of the central points Wittgenstein makes in the *Investigations* is that psychological states like meanings and intentions “seem to hover, puzzlingly and unstably, between two paradigms” (Wright, 1989b, p. 237); we seem to be drawn to conceiving them in two different, conflicting ways. On the one hand, they seem to be like conscious processes such as headaches and visual experiences, which “may have a determinate onset and departure” (Wright, 1989b, p. 237) and seem to be introspectible objects of consciousness. On the other hand, they can also seem to be more like qualities of character such as being punctual or dishonest which are dispositional and are only fully manifested in one’s behaviour. We are drawn to the first of these because of the fact that individuals have special authority about what they intend — individuals seem to have special access to the content of their own intentions, superior to that available to others. But this is problematic, Wright argues, because:

We cannot, honestly, find anything to *be* the intention, etc., when we turn our gaze inward; and anything we might find would have no connection, or at best, the wrong kind of connection with those subsequent events — what we go on to account as fulfilment of an intention or expectation, etc. — on which the correctness of an earlier ascription to us of the intention, etc., depends. (Wright, 1989b, p. 237, emphasis in original).

Not only does there not seem to be anything inner which constitutes having an intention, it’s also hard to see how anything we could find in our mental life that could have the dispositional properties of intentions and also be partly constituted by one’s later behaviour. But if we try to adopt a view of intentions which focuses on their dispositional properties then it’s hard to make sense of the special epistemic authority of the first-person perspective; why do first person avowals have special authority when determining what someone intends? If they were just like dispositions then anyone who could see their behaviour would be in an equally good position to judge what intention the speaker has in mind.

In summary, the key problem for making sense of intentions, meanings, etc., is that “their combination of first-person avowability and ‘[disposition-like]¹⁷ theoreticity’ ” (Wright, 1989b, p. 237) prevents them from fitting into either of the paradigms for other psychological states. Wright develops a judgement-dependent account of intentions which he claims can capture both these aspects.

¹⁷In “Wittgenstein’s Rule-following Considerations and the Central Project of Theoretical Linguistics” Wright refers to the latter property simply as “theoreticity”, but I find his description of it as “disposition-like theoreticity” (in “Self-Knowledge: The Wittgensteinian Legacy” (Wright, 1998/2000, p. 30)) to be much more illuminating, so I’ll use that phrase in the present chapter.

Unfortunately, the judgement-dependent account of intentions isn't quite as straightforward as that of colour. Wright (1989b, pp. 250–252) argues that to guarantee judgements about the inner are correct the C-conditions will need to involve the claim that there is no-self-deception involved, the subject is appropriately attentive, and that they have the appropriate concepts. What's problematic here is the no-self-deception condition.

The condition against self-deception is needed to take into account the possibility that the subject be deceiving themselves. Wright points out that this may be because the subject be motivated to hold a false belief about her intentions — such as in the case where she wants to believe she has formed a certain intention, but actually, because the task is hard, has no such intention — or due to some physiological or pharmacological explanation. This condition is problematic because it looks like it will almost need “a condition to the effect that the subject be ‘free of any condition which might somehow impede his ability reliably to certify his own intentions’ ” (Wright, 1989b, p. 251), which sounds very much like the kind of “whatever-it-takes” condition the non-triviality condition (ii) requires we do without.

Wright's response to this problem is to claim that the no-self-deception condition is “positive-presumptive” (Wright, 1989b, p. 251). This means that when ascribing an intention to someone we are *a priori* entitled to assume, if we have no evidence to the contrary, that the subject is not self-deceived. Now consider the C-conditions of the biconditional equation again — there is no-self-deception involved, the subject is appropriately attentive and that they have the appropriate concepts. So long as there isn't evidence to the contrary we are *a priori* entitled to assume that the no-self-deception clause is satisfied, and then we can use a restricted version of the equation, with the no-self-deception clause omitted from the C-conditions.

Let's consider an example. According to Wright it is *a priori* that:

If (x is not self-deceived, and
 x is appropriately attentive, and
 x has the concept of *intending to buy a coffee*) then,
 (x is intending to buy a coffee, if and only if x believes that she is
 intending to buy a coffee).

Now, in the case where we have no reason to believe that x may be self-deceived then, because no-self-deception is “positive-presumptive”, we are *a priori* entitled to assume that there is no self-deception going on. This means that we can omit that clause from the C-conditions, and the following biconditional will be *a priori* reasonable to believe:

If (x is appropriately attentive, and
 x has the concept of *intending to buy a coffee*) then,

(x is intending to buy a coffee, if and only if x believes that she is intending to buy a coffee).

The C-conditions of this latter biconditional do satisfy all four tests (i)–(iv) and so we should conclude that best opinions about intentions are extension-determining. But we had to factor out self-deception using its positive-presumptiveness, and so it is only *a priori reasonable* to believe the latter biconditional. This leads Wright to the conclusion that “the restricted set of C-conditions, play a *defeasible* extension-determining role” (Wright, 1989b, p. 252), which means that their satisfaction does not guarantee the truth of the subject’s judgements, but when we have no evidence that there is self-deception involved we are justified in believing their best judgements about their intentions are extension-determining.

This treatment of the self-deception clause in the C-conditions also enables Wright to account for their problematic “combination of first-person avowability and ‘[disposition-like] theoreticity’ ” (Wright, 1989b, p. 237). He does this as follows:

What determines the distribution of truth-values among ascriptions of intention to a subject who has the conceptual resources to understand those ascriptions and is attentive to them are, in the first instance, nothing but the details of the subject’s self-conception in the relevant respects. If the assignment of truth-values, so effected, generates behavioural singularities — the subject’s behaviour clashes with ingredients in his/her self-conception, or seems to call for the inclusion of ingredients which he/she is unwilling to include — then the self-deception proviso, broadly interpreted as above, may be invoked, and the subject’s opinion, or lack of it, overridden. (Wright, 1989b, p. 253).

Wright’s claim here is that we can account for the epistemic primacy of first-person avowals because, when we have no reason to doubt what they say, a subject’s best opinion constitutes their having that intention. But we can still hold on to the “disposition-like theoreticity” of intentions because if what the subject claims their intention to be conflicts with how they behave we can use that as evidence to justify the conclusion that the no-self-deception condition was in fact not met and so their opinion does not constitute having that intention. (The idea here is that if they ceased to be self-deceived, perhaps by recognising the inconsistency in their behaviour, and met the other C-conditions then they would form the correct judgement about their intention, and *that* is what constitutes their having that particular intention.)

Another detail that it’s important to note is that the judgements which constitute having a particular intention don’t need to be those that would be made, were the C-conditions to be met, at the moment the intention was originally formed. Instead, Wright claims that “subject’s best opinions about their intentions, *both past and present*, are

properly conceived as provisionally extension-determining” (Wright, 1989b, p. 254, emphasis in original). This means that when asked what constitutes the fact that a subject formed a particular intention in the past, we don’t need to turn to anything other than the fact that the subject would *now* judge themselves to have formed that intention, if the C-conditions are met.

2.5.4 The judgement-dependent account of meaning

Wright argues that we can adopt a parallel account of meaning, and this provides us with just what we need to respond the problems posed by Kripke:

It will be . . . a perfect answer to Kripke’s Sceptic to explain how judgements concerning one’s own meanings, both past and present, are likewise provisionally extension-determining in the most ordinary circumstances. Challenged to justify the claim that I formerly meant addition by ‘plus’, it will not be necessary to locate some meaning-constitutive fact in my former behaviour or mental life. A sufficient answer need only advert to my present opinion, that addition is what I formerly meant, and still mean, and to the *a priori* reasonableness of the supposition, failing evidence to the contrary, that this opinion is best.

Responding to Kripke’s Sceptic in this way does not require construal of meaning as a kind of intention; it is enough that the concepts are relevantly similar — that both sustain authoritative first-person avowals, and that this circumstance is to be explained in terms of failure of the order-of-determination test. (Wright, 1989b, p. 254).

This explains why knowledge of our meaning is non-inferential, if knowledge of what we mean isn’t based on detecting some fact independent of our judgements, but is instead constituted by our best judgements then we should not expect to find some independent fact from which we infer knowledge of what we mean. And when responding to Kripke’s sceptical challenge as to how we know that what we meant by ‘plus’ yesterday is the same thing we currently mean, we can simply respond by referring to our current judgement that the meanings are the same. Our best judgements play a defeasible extension determining role, so “failing evidence to the contrary, that opinion is always best” (Wright, 1989b, p. 254); nothing more is needed to constitute the fact that I mean the same thing by ‘plus’ today as I did yesterday.

This solution is also not vulnerable to the criticisms which we saw Boghossian and McDowell raising against Wright’s earlier view (from *Wittgenstein on the Foundations of Mathematics* (Wright, 1980)). As we saw in §2.4.2 Boghossian demonstrated that Wright’s earlier response to rule-following got the extension of words wrong because it would come out with the conclusion that ‘horse’ actually meant *horse or horsey-looking-cow-at-night or . . .* . This happened because, as McDowell (1984/1998o, p. 234) argued, Wright couldn’t make sense of the distinction between saying ‘This is yellow’ and ‘This would be called

‘yellow’ by (most) speakers of English’. The judgement-dependent account can respond to both of these problems because the community’s judgements in the C-conditions and their actual judgements can come apart:

it is a perfectly objective question what, in a particular case, the deliverance of best opinion would be; and that deliverance is something with which a majority, or even a whole community, may for some reason be out of accord. (Wright, 1989b, p. 249).

2.6 Evaluation of Wright’s judgement-dependent account

In “Intentionality and Interiority in Wittgenstein” McDowell (1991/1998f) argues that the role subsequent judgements play in Wright’s judgement-dependent account of meaning conflicts with certain platitudes about meaning. Wright responds by arguing that McDowell is guilty of mischaracterising his position, and although I somewhat agree with Wright here looking at this dispute is valuable because it reveals a deeper problem in Wright’s account.

According to McDowell, it is Wright’s view that what particular meaning you have in mind (say, whether you mean *plus* or *quus* by ‘+’) is dependent upon your subsequent judgements on the matter. If you go on give the answer ‘5’ when asked ‘What’s $68 + 57$?’ and judge that you are acting in accord with your previous use of ‘+’ then that means that by ‘+’ you actually meant *quus*, even in your past usage. McDowell attacks this view as follows:

suppose I form the intention to type a period. If that is my intention, it is settled that only my typing a period will count as executing it ... if that is the intention that ... I form, nothing more than the intention itself is needed to determine what counts as conformity to it. Certainly it needs no help from my subsequent judgements. (Suppose I forget what a period is.) So there is something for my intention to type a period, conceived as determining what counts as conformity to it autonomously and independently of my judgments on the matter, to be; namely, precisely, my intention to type a period. An intention to type a period is exactly something that must be conceived in this way. ... [It is] common sense [that] an intention is exactly something that can be, in some sense, all there in one’s mind before one acts on it. One does not need to wait and see what one does before one can know what one intends. (McDowell, 1991/1998f, p. 315).

So, McDowell’s argument here is that Wright’s judgement-dependent account entails that the content of your intentions is dependent upon what you happen to go on to do. McDowell criticises this, arguing that it is common sense that all that’s needed for an intention to have a determinate content is for someone to form the intention — it is “all there in one’s mind before one acts on it” — and we do not need to “wait and see what one does”

in order to know the content of an intention. He argues that our subsequent judgements may go wrong because a subject may, for example, forget what a period is. In that case then they will not judge that they formerly intended to type a period, and so their current judgements clearly can't constitute what it is for them to have formed that intention in the past.¹⁸ Wright's view completely fails to connect with our ordinary notion of intentions, and so must be rejected.¹⁹

At first sight, McDowell's criticism of Wright seems fitting, at least when we look at what Wright says in "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention":

But perhaps the most decisive difficulty that his [Wittgenstein's] writings disclose [in the Cartesian conception of knowledge of one's own intentional states] derives from the requirement the model imposes that intentional states can be fully determinate objects of inner contemplation before they issue in anything outward — and hence that their connection with events, and especially performances, lying in the future can only be causal. For, on the contrary, the ascription of intentional states to a subject is answerable to what he goes on to say and do in the broadly identificatory fashion in which the ascription of dispositions and capacities is so answerable. (Wright, 1989a, p. 631).

Here Wright seems to be endorsing what he takes to be Wittgenstein's criticism of the Cartesian picture for claiming that intentions can be determinate before they have any behavioural manifestation. Indeed, Wright claims that ascriptions of intentions are subject to the individual's future behaviour. This is precisely what McDowell is accusing him of.

However, when he responds to McDowell's criticism, in "Self-Knowledge: The Wittgensteinian Legacy", Wright seems to disagree with this characterisation of his view:

Of course, if I form the intention that P, what will comply with it is only and exactly the bringing it about that P; and it will typically be a matter independent of my subsequent judgements whether or not just that has been effected. The role of subsequent judgements is indeed not to mediate somehow in the connection between the content of an intention and its execution (Wright, 1998/2000, p. 29).

¹⁸This particular example of forgetting what a period was isn't very helpful since one of the C-conditions is that the subject has the relevant concepts. And so the speaker's judgements in that case will be irrelevant in determining the content of the intention. However, as we'll see in a few pages, this example is easily modified to provide one which is more problematic for Wright.

¹⁹In the paper "Are Meaning, Understanding, etc, Definite States?" McDowell (2009a, p. 99–95) accuses David Pears of making the same mistake. He points out that Pears claims that when one first claims to understand a word that offers only "incomplete guidance" which needs to be "supplemented . . . by the rule-follower's propensity to find it natural to write one number rather than another at the relevant point in the series" (McDowell, 2009a, p. 93). In a footnote he claims that Wright is making a similar mistake (McDowell, 2009a, p. 93, fn. 10).

And this seems to fit well with Wright's view. As we saw above, he thinks that:

If (x is not self-deceived, and
 x is appropriately attentive, and
 x has the concept of *intending to P*) then,
(x is intending to P , if and only if x believes that she is intending to P).

When a subject initially forms the intention what makes it the case that she is forming that particular intention is that it is the intention she would believe herself to have formed if the C-conditions are met. There seems to be no need for this account to involve subsequent judgements, instead all that matters is what she currently believes herself to intend (if the C-conditions are met). And so, McDowell is wrong to claim that Wright is committed to the idea that an intention needs help from subsequent judgements in order to have a determinate content.

So, what role do subsequent judgements play in Wright's account? After the passage quoted above, Wright continues:

[The role of subsequent judgements is] to enter into the determination of *what* the content of an *anterior intention* is to be understood as having been. . . . there is nothing for my intention's having had *just that content* to consist in, if the fact has to be constitutively independent of anything which I may subsequently have to say about compliance or non-compliance with the intention, or what it's content was. (Wright, 1998/2000, p. 29, some emphasis added).

The idea here seems to be that when judging what the content of an earlier intention was we can make use of evidence from the subject's behaviour after forming an intention. If someone believes themselves to have formed the intention to stop lying, but then continues to regularly lie then it would be reasonable for them to conclude that they may not have sincerely formed that intention in the first place. In that case Wright would argue that the no-self-deception C-condition was not in fact met, so their original belief that they intended to be honest does not show that they actually had that intention. As we saw above, this aspect of Wright's view enables him to account for the "disposition-like theoreticity" of intentions, the fact that judgements about intentions are partly answerable to the subject's future behaviour.

I think that this aspect of Wright's account is deeply problematic. In order to get clarity on what's going on here I'm going to draw on Edwards' explanation of Wright's view in "Best Opinion and Intentional States":

Wright's idea is that just as his original judgement that he had that particular intention constitutively determined that he did indeed have that intention,

so his later judgement, that writing ‘14, 16, 18’ implemented rather than frustrated that intention, constitutes the fact that such behaviour does implement that particular intention. A subject’s opinion as to whether or not he has changed his mind, abandoned his former intention, determines whether or not he has changed his mind. So best opinion as to our current intentions determines those intentions, and best opinion as to what current behaviour implements a former intention determines what behaviour implements that former intention. And this determination in each case is not causal or contingent. Rather, his best opinion as to what his current intentional state is, and his future opinion as to what behaviour his former intentional state now requires, *canonically individuate* that state (Edwards, 1992, p. 24, emphasis in original).

Most usefully to our present concern, Edwards formalises Wright’s account of past and future intentions as follows:

Wright takes the schemas

(3) $Cx \rightarrow (x \text{ believes } x \text{ intends } P \leftrightarrow x \text{ intends } P)$

(4) $Cx \rightarrow (x \text{ believes } x \text{ intended } P \leftrightarrow x \text{ intended } P)$

to be *a priori* and *constitutive* of our concept of intention, when ‘ Cx ’ is defined non-trivially. (Edwards, 1992, p. 24).²⁰

Edwards is not critical about this aspect of Wright’s account, but I think it reveals a deep problem with Wright’s view, in that what determines the content of a particular intention seems to change over time. Consider McDowell’s example of someone intending to type a period. If a person forms that intention at 3pm then what makes it the case that that is the intention she has formed is that, were the C-conditions to be satisfied, she would believe that is the intention she has formed. Indeed, what *constitutes* her having the intention at that time is the disposition to form that belief.

But now let’s consider the same person at 4pm. According to Wright’s account what now makes it the case that she formed the intention to type a period is the fact that, if the C-conditions are satisfied, she *now* believes that was the content of her intention. And now *that* constitutes what it is for her to have had the intention an hour ago. So, what constitutes her forming the intention to type a period at 3pm changes over time, it is always constituted by her *present* beliefs. To put it more formally, what this means is:

At T1: If (Cx at T1) then, (x is intending to P at T1), if and only if (at T1 x believes that she is intending to P at T1).

At T2: If (Cx at T2) then, (at T2 x intended to P at T1), if and only if (at T2 x believes that she intended to P at T1).

²⁰ “ Cx ” at the start of these schemas refers to the conditions in which the C-conditions are satisfied, i.e. the conditions when she is not self-deceived, and is appropriately attentive, and has the relevant concepts.

At T1 and T2 different things constitute the very same intention (the one which was formed at T1): at T2 her beliefs at T2 are relevant, whereas at T1 only her beliefs at T1 are relevant.

So, what if x 's beliefs about her intention change? Well, Wright attempted to deal with this possibility by claiming that we should then invoke the no-self-deception clause in the C-conditions:

If the assignment of truth-values, so effected, generates behavioural singularities — the subject's behaviour clashes with ingredients in his/her self-conception, or seems to call for the inclusion of ingredients which he/she is unwilling to include — then the self-deception proviso, broadly interpreted as above, may be invoked, and the subject's [original] opinion, or lack of it, overridden. (Wright, 1989b, p. 253).²¹

This seems appropriate for the case of the person who believed herself to have formed the intention to give up lying, which I considered above. In that case her future continued deception would lead her to doubt that she had ever genuinely formed that intention, and conclude that her original belief in her honesty was a result of self-deception. This means that at that earlier time the C-conditions were in fact not met, and so we can disregard her original belief in her intention to be honest. And this was a good result for Wright, since it enabled him to make sense of the “disposition-like theoreticity” of intentions.

But merely fitting with a few examples is not enough. What Wright needs for his position to avoid the apparent conflict I identified is the claim that there can be no case where the C-conditions are met at two different times, and yet the subject changes her beliefs about what intention she had at a particular time. He needs there to be an *a priori* guarantee that whenever a speaker changes her mind about the content of a past intention then when she formed one of the beliefs the C-conditions were not satisfied. If he had this then his account would be successful; judgements made in C-conditions constitute the content of a subject's intentions, and he's accounted for both the first-person authority of intentions and their “disposition-like theoreticity”. The problem is that Wright presents us with no evidence for this claim, and merely assuming it seems completely *ad hoc*.

Indeed, not only does the assumption of this principle seem to be *ad hoc*, in actual fact it appears to be false: there are cases in which a speaker's belief about her past intention can change and the C-conditions be satisfied when she formed both of the conflicting beliefs.

As we saw above McDowell attempts to develop such an example by pointing out that a speaker may “forget what a period is” (McDowell, 1991/1998f, p. 315). If a speaker forgets what a period is then clearly at T2 she won't believe that at T1 she formed the

²¹I added the word ‘original’ to this quotation to make it clear which opinion is relevant. In the original passage Wright could be talking about other people's opinions overriding the subject's opinion, but that is not significant for the point I am making.

intention to type a period. However, this example is no help since one of the C-conditions is that the speaker has the relevant concepts, and if she has forgotten what a period is then clearly that won't be the case. Wright can simply discount her latter judgements about whether or not she had the intention to type a period, since at that time the C-conditions have not been satisfied.

But what about the more straightforward case where a speaker simply forgets what she intended; she still has full understanding of the concept of a period, she's simply forgotten what she intended, and so now would not believe that she intended to type a period? This is a fairly common case, people forget what their intentions are all the time, even sometimes before they've actually carried them out! (Who hasn't walked into a room with a specific intention only to arrive in the room having forgotten what they went in there for in the first place?) In such a case there seems to be no reason to claim that the C-conditions can't possibly continue to be satisfied. People can forget their intentions without any self-deception involved, or without failing to properly attend to their intentions. Let's say that at T1 she satisfied the C-conditions and believed herself to be forming the intention to type a period. And at T2 she still satisfied the C-conditions, but has simply forgotten her intention, and so has no disposition to form the belief that she intended to type a period. According to Wright's account this means that at T1 it is true that at T1 she intends to type a period; whereas at T2 it is false that at T1 she intended to type a period! Clearly something has gone wrong!

Perhaps the reader will not find this particular example to be very convincing, but that's not key here, all we need for this to undermine Wright's account is his failing to provide an *a priori* guarantee that this sort of case is impossible.

In response to this problem, if he cannot find such an *a priori* guarantee then Wright seems to have two options, both of which fail to fit with one of the two paradigms that he claims intentions hover between, and so fail to fully capture our ordinary conception of intentions.

On the one hand Wright could prioritise the latter judgement. In actual fact in "Self-Knowledge: The Wittgensteinian Legacy" Wright (1998/2000) *seems*²² to be doing

²²I say "seems" here because it's hard to develop a consistent reading of this part of Wright's paper. A little higher up the same page, in a passage which I've already quoted, he seems to say precisely the opposite:

Of course, if I form the intention that P, what will comply with it is only and exactly the bringing it about that P; and it will typically be a matter independent of my subsequent judgements whether or not just that has been effected. The role of subsequent judgements is indeed not to mediate somehow in the connection between the content of an intention and its execution (Wright, 1998/2000, p. 29).

precisely this:

the content of a subject's intentional states . . . is something which is intrinsically sensitive to the deliverances of best interpretive methodology. That is a methodology which in principle must include within its conspectus the whole sweep of a subject's sayings and doing, including future ones, without bound. (Wright, 1998/2000, p. 29).

The problem with this option is that it doesn't account for the properties that intentions share with conscious processes like headaches and visual experiences — their first-person avowability. As we saw above, according to Wright we can have introspective access to the content of our intentions, but if their content depends upon what our future behaviour happens to be, then this would seem to be impossible.

The alternative to this solution would be to prioritise the original judgement. This would mean that we would reject the claim that a speaker's future judgements can determine the content of a past intention. And so we would have to reject the claim that:

At T2: If (Cx at T2) then, (at T2 x intended to P at T1), if and only if (at T2 x believes that she intended to P at T1).

And modify the basic judgement-dependent formula to be something along the lines of:

If (x is not self-deceived, and
 x is appropriately attentive, and
 x has the concept of *intending to P*) then,
(x is *currently* intending to P , if and only if x *currently* believes that she is intending to P).

This option fails because it conflicts with the other of the two paradigms that Wright claims that intentions hover between — their “disposition-like theoreticity”. According to Wright the content of an intention is partly constituted by one's later behaviour, and if we make this move then we'll completely break the constitutive link between a speaker's intention and their future behaviour.²³

Wright is trapped. Either he needs to provide us with an *a priori* guarantee that if a subject changes her mind about what she intended at a particular time then when she held one of the conflicting beliefs the C-conditions must not have been satisfied, or he will be forced to emphasise either the earlier or the later judgement, both of which end up conflicting with what he claims to be fundamental properties of intentions. Unless he can provide us with the required *a priori* guarantee (and it seems counter-intuitive, so unlikely) a judgement-dependent account of intentions seems hopeless.

²³It could conceivably be argued that these are not the only two options and that in fact there could be an account which sometimes favoured the judgement made at T1, and other times favoured the judgement made at T2. However, we would then need a principled reason to favour a particular time's judgement in each case, and it's not at all clear to me how such a solution would work.

Chapter 3

Rule-following part 2: McDowell's solution

In *Wittgenstein on Rules and Private Language* Kripke (1982) challenged us to find a fact which constitutes our grasp of meanings. Any response to this challenge needs to do two things: it must determine the extension of our expressions, even for cases so far unconsidered, and it must explain why we *should* use a word in a particular way — it must be normative. Kripke argued that there is no such fact, but his own “sceptical solution” ended in disaster, so it seems we should instead aim to find a fact that satisfies Kripke’s two requirements.

Another thing we learnt from Kripke was how essential it was for any account of meaning to be able to draw the seems right / is right distinction — a successful account needs to be able to account for the fact that we sometimes make mistakes in our judgements as to how to use language. Wright’s early attempt at a solution to the rule-following problem in *Wittgenstein on the Foundations of Mathematics* (Wright, 1980) accounted for this at the level of individuals, but couldn’t make sense of whole communities making such mistakes.

Wright’s judgement-dependent account of meaning successfully accounted for the fact that we will sometimes be disposed to make mistakes in our use of language by claiming that the facts about what we mean consist only in our judgements made in best conditions. And in claiming that best judgements *constitute* the facts about meaning this account is also able to provide a fact which is able to satisfy Kripke’s two demands. The reason Kripke failed to find such a fact was that he was assuming that we could only have inferential knowledge of our meanings, and so would be able to infer what we mean from our past behaviour and mental life. Wright’s judgement-dependent account enables us to make sense of the idea that we can have non-inferential knowledge of such facts by developing a non-reductionist account of meaning. However, as we saw at the end of the last chapter, this solution is problematic because it seems that we can change our

mind about what we meant, despite continuing to satisfy the conditions required for our judgements to constitute the facts about our meaning. This leaves us with two different, potentially conflicting judgements with seeming equal claim to constitute the same fact. And so Wright's judgement-dependent account also fails to provide an adequate response to Kripke's sceptical challenge.

In this chapter I shall turn to look at McDowell's solution to the rule-following problem. McDowell, like Wright in his later account, endorses a non-reductionist account of meaning and claims that we have non-inferential knowledge of what we mean, which is why Kripke's sceptical challenge was so difficult to respond to. However, he attempts to hold this position without claiming that best judgements constitute meanings, and so avoids the problem which I pointed out at the end of last chapter. Instead, he develops a quietist solution, which attempts to remind us of common sense facts in order to reveal that we shouldn't be worried how it is possible for us to have non-inferential knowledge of what we mean.

This view may initially seem problematic, since it looks like McDowell's response to the question of how meaning is possible is simply to assert that it is! In response to this worry I shall explain McDowell's view of our knowledge of what we mean in the context of his overall epistemological picture. I'll go through McDowell's account of our experience of the world, and briefly look at his justification for this view. Although I won't attempt to evaluate the epistemological assumptions which justify McDowell's view, I hope to at least make clear how his view of meaning fits into his overall view of the world and our experience of it.

Finally I shall look at McDowell's response to the seems right / is right distinction. McDowell argues that we can make sense of this distinction by pointing to the fact that we can consider what we meant from second and third-person points of view, and recognise when someone's made a mistake. However, I will argue that other things he says about the inner world prevent him from being able to draw the seems right / is right distinction in all the cases in which he should. And so I shall conclude that although his response to the rule-following problem succeeds, his account of the inner is problematic.

I shall begin my exegesis of McDowell by looking at how he thinks Wright goes wrong. But first it will be useful to have a brief summary of our conclusions from the previous chapter.

Kripke

A straight solution must provide a fact which constitutes having a particular meaning in mind that which satisfies the two constraints: it must be able to determine the correct

extension of the relevant expressions, and it must be normative — it explains how speakers know they should use words in a particular way.

Alternatively we can go for a sceptical solution, which accepts that there is no such fact, but still needs to be able to draw the seems right / is right distinction (otherwise whatever seems right would be right, and we couldn't make sense of the idea that we are following rules at all).

Kripke attempted a sceptical solution which accepted that statements about meaning don't have truth-conditions, and instead relied on their assertion conditions. This failed because Kripke's own arguments can be used to show that there are no facts about the assertion conditions either. So, it would be ideal if we could find a straight solution to the sceptical challenge, and find a fact which constitutes having a meaning in mind and passes Kripke's two tests.

Wright's communitarian account

The facts that constitute meaning reside in the community's dispositions to use language.

This view fails because the community has dispositions to make mistakes, and sometimes to consistently make the same mistakes (such as using 'horse' to refer to horsey-looking cows on dark nights). This shows us we must be able to draw the seems right / is right distinction at the level of whole communities, not just individuals.

Wright's judgement-dependent account

Perhaps meanings are non-reductive? So we could have non-inferential knowledge of what we mean, and that explains why we couldn't respond to Kripke's sceptical challenge because he's asking us to give an account of the facts that constitute meaning without referring to facts about meaning themselves.

Wright claims we need to flesh out such an account of meaning, and so develops a judgement-dependent account — what you mean is constituted by what you would judge yourself to mean in appropriate conditions.

The problem with this account is that Wright is committed to different judgements (those made at different times) constituting the same fact about what you meant at a particular time, and it seems that these could potentially conflict.

Summary: conclusions about straight solutions to the problem of meaning

- We need to give an account of a fact that constitutes having a meaning in mind which is normatively significant and determines the correct extension of our expressions.
- We need to be able to draw the seems right / is right distinction, both for individuals and for whole communities.

- Meaning could be non-reductive; the fact which constitutes meaning could be known non-inferentially.
- If meaning is non-reductive then we need some kind of account of it, and judgement-dependent accounts seem to fail.

3.1 McDowell’s diagnosis of Wright’s mistake

McDowell, like Wright,¹ rejects reductionism about meaning, but his solution to the rule-following problem differs from Wright’s because he doesn’t think that to endorse a non-reductionist account of meaning we need to introduce anything like judgement-dependence. Before getting into the details of McDowell’s solution I shall explain why he thinks Wright goes wrong, so we can best make sense of McDowell’s project.

As we saw in §2.5.3, Wright argues that one goal of Wittgenstein’s arguments about rule-following was to demonstrate their “disposition-like theoreticity” (Wright, 1989b, p. 237). He rejects the Cartesian view that we know what meaning we have in mind simply by awareness of an inner world — he thinks that “we do not cognitively interact with states of affairs which confer truth upon our opinions concerning our own intentional states” (Wright, 1989a, pp. 632–633). This leads Wright to completely reject the idea that we know our meanings by awareness of inner states, and indeed to drop the idea that there are any facts about meaning which are independent of our best judgements.

McDowell thinks that Wright takes Wittgenstein’s criticism of the Cartesian view too far. He agrees that the Cartesian account of how awareness of an inner world enables knowledge of our meaning is deeply flawed, but instead of completely abandoning talk of inner awareness he proposes an alternative account of the nature of inner awareness which aims to avoid Descartes’ mistakes, and models our inner awareness on his view of our experience of the outer world. Before getting into that view I’m first going to explain his account of Wright’s mistake.

According to McDowell:

In Wright’s reading, Wittgenstein in effect [denies] that understanding *could* be a matter of having something in mind; the idea of having something in mind fits the other main compartment of our subject-matter, the occurrent phenomena of consciousness, and they do not involve intentionality. But now, if our location in a “normatively” shaped space is not constituted by having meanings in mind ... then what does constitute it? (McDowell, 1991/1998f, p. 305, emphasis in original).

¹In this chapter unless otherwise stated, whenever I talk about “Wright” I mean to refer to Wright’s view when he endorsed a judgement-dependent solution to rule-following.

Instead of meaning involving simply having something in mind Wright claims that our judgements about what we mean, when made in the appropriate conditions, themselves *constitute* meaning. This enables him to explain how there can be facts about what we mean which don't reduce to some other fact (the lack of which was why Kripke's sceptical challenge was so hard to answer) without resorting to mysterious talk about awareness of an inner world.

Wright claims that Wittgenstein's official solution to the rule-following problem is to undermine the Cartesian picture of the inner but give us nothing in its place and instead just to urge quietism:

Asked what constitutes the truth of rule-informed judgements ... the official Wittgensteinian will reply: 'Bad question, leading to bad philosophy — platonism, for instance, or Kripkean scepticism.' (Wright, 1989b, p. 257).

Wright agrees with Wittgenstein that an adequate account of meaning doesn't require us to reduce meanings to something else (the mistake at the heart of Kripke's sceptical challenge was to assume reductionism). But he's deeply unhappy with Wittgenstein's quietism and thinks something along the lines of his own judgement-dependent account is essential for a philosophically satisfying explanation of how it is possible for meanings to come to mind.

But, as we saw above, Wright's judgement-dependent account of meaning doesn't work. Or, at least, it conflicts with deeply held intuitions about the nature of meaning and intentions, intuitions which Wright himself thinks any successful account of meaning must be compatible with.

Unlike Wright, McDowell does not think that Wittgenstein's refusal to answer certain questions about how meaning is possible makes him guilty of a philosophically unsatisfying quietism. Instead, he attempts to develop an account of awareness of the inner which he claims dissolves the need for an answer to the "how possible" question, by arguing that inner awareness (like our outer experience of the world) is conceptually structured.²

3.2 The master thesis

Central to McDowell's response to the rule-following problem is the rejection of an assumption which he calls "the master thesis":

the thrust of Wittgenstein's reflections is to cast doubt on the master thesis:
the thesis that whatever a person has in her mind, it is only by virtue of

²I'll explain in detail what McDowell means by "conceptually structured" in §3.3.1, but the key idea in claiming that experience is conceptually structured is that the perceivers' conceptual capacities are involved in their experience. This means that experience of a blue object would not be simply a brute experience, but would involve experience of the (conceptually structured) fact that *that object is blue* — which necessarily involves the concept *blue*.

being interpreted, in one of various possible ways, that it can impose a sorting of extra-mental items into those that accord with it and those that do not. (McDowell, 1993/1998h, p. 270).

McDowell equates rejecting the master thesis with endorsing Wittgenstein's claim that:

there is a way of grasping a rule which is *not* an *interpretation* (Wittgenstein, 1953/2001, §201, emphasis in original).

We saw in §2.1.2 the example of interpretations of mental images. These are rules for telling you how to apply the word associated with a particular mental image — rules which give the normative significance of the mental image. (In §2.1.2 we saw the sample of the rule of interpretation “ ‘Triangle’ applies to x if and only if x looks like \triangle ”, this tells us how to interpret the mental image \triangle in order to know how to apply the word ‘triangle’.) But interpretations don't apply only to mental images, instead, an interpretation is a rule for telling us how to correctly apply anything, such as mental images, objects, or even rules themselves. Interpretations take something normatively inert and give it normative significance.

McDowell argues that if we assume that the master thesis is true — things in mind need to be interpreted to have normative significance — then we will be forced into a view along the lines of Kripke's:

Kripke's reading of how the regress of interpretations threatens the very idea of understanding turns on this thesis: “no matter what is in my mind at a given time, I am free in the future to interpret it in different ways” [(Kripke, 1982, p. 107)]. This presupposes that whatever is in a person's mind at any time, it needs interpretation if it is to sort items outside the mind into those that are in accord with it and those that are not. There are always other possible interpretations, and a different interpretation, imposing a different sorting, may be adopted at a different time. Considered in themselves, that is, in abstraction from any interpretations, things in the mind just “stand there” (McDowell, 1993/1998h, p. 268).

We saw at the very beginning of this chapter that according to Kripke any successful account of the facts that constitute meaning must be able to explain the normativity of meaning. For example, we need to be able to explain why it is that when asked ‘What's $68 + 57$?’ we *should* answer ‘125’ and not ‘5’, and why the word ‘cow’ applies to cows, but not horses — the facts which constitute meaning need to be able to “sort items outside the mind into those that are in accord with it and those that are not” (McDowell, 1993/1998h, p. 268). McDowell argues that Kripke's sceptical argument involves the assumption that what is in the mind needs interpretation in order to have any normative force; this assumption is essential for Kripke's worries to get off the ground and that if we reject it then his sceptical argument collapses.

If we accept that things in the mind need interpretation to be normative then we get the problems which we saw in §2.1.2. At first sight, it seemed that we could simply use a mental item (such as a mental image of a triangle) to specify the normative significance of a word. But Kripke pointed out that for any mental item we also need to know *how* to interpret it, because for anything which comes to mind there will always be alternative rules of interpretation which fit equally well with our past behaviour and mental life. This was made clear with the example where we attempted to use the mental image of a triangle in order to interpret the word ‘triangle’. In that case it initially seemed that we could simply use the rule of interpretation:

‘Triangle’ applies to x if and only if x looks like \triangle .

to use a mental image of a triangle to know when we should and shouldn’t use the word ‘triangle’ — to give ‘triangle’ its correct normative significance. The problem here is that this rule alone isn’t sufficient to give us the meaning, since we need a further rule of interpretation to know the meaning of ‘looks like’ here, and there are alternative understandings of ‘looks like’, other than *looks-like*, such as:

x *quooks like* \triangle if and only if
 ((the year is 2013 or earlier and x looks like \triangle), or
 (the year is greater than 2013 and x has one side more than \triangle)).

Kripke’s sceptic would argue that there is nothing in our past behaviour, or mental life, to indicate whether we understand ‘looks like’ in the first rule of interpretation to mean *looks like* or *quooks like*. We could follow though with the same strategy we used to fix our understanding of ‘triangle’ and use a further mental image to know how to interpret ‘looks like’ and ensure we avoid deviant meanings such as *quooks like*. But for that mental image the same problem will occur, it too will require interpretation in order to have a determinate meaning — “Once again, a rule for interpreting a rule.” (Kripke, 1982, p. 43) — and we are led into an infinite regress.

It seems that when attempting to interpret something we will always have to use a further rule of interpretation to give it the correct normative force. But if that further rule of interpretation requires interpretation itself to have a determinate meaning then we’ll be led into an infinite regress which can never be worked though.

McDowell argues that the right response to this problem is not to endorse scepticism about meaning (and, as we saw in chapter 2, such a view ends in disaster — Kripke’s attempt to claim that there are no facts that constitute meaning was ultimately found unable to draw the seems right / is right distinction, and lost grip of the very idea that we’re following rules at all), but instead to reject the assumption which enables the sceptical worries to get going; we need to recognise that “there is a way of grasping a rule which is *not* an *interpretation*” (Wittgenstein, 1953/2001, p. 229, emphasis in original).

In other words, we need to reject the assumption that mental items require interpretation to have normative significance, and so can avoid the regress.

In addition to arguing that the assumption that grasping a rule must involve interpretation leads in disaster, McDowell also claims that endorsement of the master thesis is in fact highly counterintuitive:

Suppose I am struck by the thought that *people are talking about me in the next room*. The supposition implies that only a state of affairs in which people are talking about me in the next room would be in accord with my thought. Now the master thesis implies that whatever I have in mind on this occasion, it cannot be something to whose very identity that normative link to the objective world is essential. It is at most something that can be interpreted in a way that introduces that normative link, although it can also be interpreted differently. . . . Considered in itself it has no relations of accord or conflict to matters outside my own mind, but just “stands there”. . . .

. . . The master thesis implies that what a person has in mind, strictly speaking, is never, say, *that people are talking about her in the next room* but at most something that *can* be interpreted as having that content, although it need not. Once we realize that that is what the master thesis implies, it should stand revealed as quite counter-intuitive, not something on which a supposed need for constructive philosophy could be convincingly based. (McDowell, 1993/1998h, pp. 270–272, emphasis in original).

McDowell here is attempting to demonstrate the strangeness of endorsing the master thesis by pointing out that if it were true then nothing in mind would of itself have a meaning (or, indeed, any kind of normative properties, such that they apply in some cases and not in others — not only could meanings not come to mind, neither could intentions or rules). Instead, everything in mind would require interpreting to be meaningful (or to be an intention, or to be a rule, or . . .).

To make vivid how strange this is McDowell reminds us of Wittgenstein’s example (Wittgenstein, 1953/2001, p. 54, fn. a) of the mental image of an old man walking up a steep path leaning on a stick. If it needed interpretation to represent some particular state of affairs (i.e. that of an old man walking up a steep path leaning on a stick) then that very same image could also be interpreted as an old man sliding downhill in that position, and so its meaning would be indeterminate. If the master thesis is true then that would mean that all mental content would be like this, in that it requires interpretation in order to have determinate normative force. Surely, McDowell insists, it would be much more commonsensical to simply hold that things can come to mind which don’t require interpreting to have a particular normative significance. Why can’t meanings and intentions themselves come to mind, then we can avoid the entire problem of things in the

mind needing interpretation to have meaning, and the devastating indeterminacy which that leads to.

McDowell's move here has, at first sight, close parallels to Wright's rejection of reductionism about meaning which we saw in §2.5.1 ("Wright's attack of Kripke's inferential assumption"). Remember that Wright accused Kripke of simply presuming that if meanings are capable of coming to mind then we must be able to reduce meanings to something else. Wright pointed out that not all knowledge is like this by using the example of our knowledge of former perceptions, which is also non-inferential. McDowell's insistence that we must be able to know what we mean without having to interpret something in mind looks just like Wright's insistence that we shouldn't simply presume that the facts about what we mean must be inferred from something else in our head.

Where McDowell and Wright differ is in how they make sense of this conclusion. Wright attempts to explain this possibility by arguing that what constitutes having a particular meaning in mind is our judgements, when made in appropriate conditions, that that is the meaning we have in mind. McDowell develops no such account, and instead seems happy to simply assert that we are able to know what we mean because things can come to mind which have normative force. If we insist that this is possible without interpretation, then we escape the regress of interpretations problem.

3.3 Conceptually structured experience

As I've explained it so far McDowell's solution to the problem appears highly unsatisfying. It almost looks as if he's just insisting that the problematic thing is possible — that mental items can have normative significance — and simply left it at that. Compared to Wright's sophisticated judgement-dependent account of meaning McDowell's seems woefully inadequate.

I think that these problems can be resolved by putting McDowell's view about meaning and awareness of what we mean in the context of his general view about the nature of our experience of the world. His arguments about the nature of the world and our experience of it contain lots of complicated and controversial elements, and attempting to evaluate the argument's success could easily fill an entire chapter itself. Instead, in this section I aim merely to explain how his argument works so that it's clear how his view of meaning links up to his overall view of our relationship with reality.³

³It may appear that the development of all this philosophical machinery conflicts with McDowell's endorsement of philosophical quietism. In §3.6 I shall look into this problem, and tentatively conclude that in fact his work here is compatible with quietism.

McDowell's view here is essentially a development of Kant's⁴ and centres around the claim that:

the world . . . cannot be constitutively independent of the space of concepts, the space where subjectivity has its being (McDowell, 1991/1998f, p. 306).

Essentially what McDowell argues for is the claim that it is not only our talk and thought which is conceptually structured, but so is the world itself and our experience of it. I'll look more closely at what McDowell means by claiming that the world is conceptually structured in §3.3.1, but the fundamental idea here is that the world consists of facts, not things. Facts are always facts *that* such-and-such is the case; facts necessarily implicate concepts "inside a content-specifying "that"-clause" (McDowell, 1987/1998e, p. 91). For example, consider the fact *that spring has begun*, this fact implicates the concepts *spring* and *beginning*. If the world is made of facts, then this means that the world itself is not constitutively independent of such concepts. In addition, McDowell argues that our experience of the world must also be conceptually structured, and ultimately his solution to the rule-following problem rests on making the very same claim about our experience of the inner world — that our experience of inner states is also conceptually structured. But I'll go into this claim in much greater detail in §3.4 below.

McDowell's argument for the claim that the world and our experience of it must be conceptually structured starts with an epistemic worry about how our beliefs about the world could be justified if the world were not conceptually structured. Although he tends to put this point in epistemic terms, McDowell thinks this issue actually goes beyond the epistemic, since if we can't explain how the world is able to "exert a rational constraint on our thinking" (McDowell, 2000/2009c, p. 42) — if we can't explain how the world is able to cause and justify our beliefs about it — then we can't even make sense of the idea that we are able to refer to the world in our thought or talk at all. He claims that the world must be conceptually structured, and believing otherwise would mean we have "obliterated empirical content altogether" (McDowell, 1994/1996, p. 43).

The crucial driving force behind McDowell's argument here is the idea that:

the world itself must exert a rational constraint on our thinking (McDowell, 1994/1996, p. 42).

Our beliefs about the world must be capable of having a rational relation to the world itself, such that when all goes well and we have knowledge of the world that is because our

⁴In actual fact, McDowell's view is closer to Hegel's, since he follows Hegel in attempting to purge the Kantian view of the idea of the noumenal, but such details aren't particularly significant here, see "Intentionality and Interiority in Wittgenstein" (McDowell, 1991/1998f, pp. 306–307).

beliefs accurately depict the world and have the right epistemological relationship with it. This links up with Sellars' claim that:

in characterising an episode or state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says (Sellars, 1956/1997, §36, emphasis in original).

The idea of the “space of reasons” is key to Sellars' and McDowell's view of knowledge, and is essentially just the name for the space of things which can be involved in justifying our beliefs. So, if I form the belief *that there is a chair*, then the chair itself must somehow be within the “space of reasons” if it is to play a role in providing me with a reason to have that belief. In saying that “the world itself must exert a rational constraint on our thinking” (McDowell, 1994/1996, p. 42), McDowell is effectively saying that the world itself must be within the space of reasons.

McDowell's argument for the claim that the world is conceptually structured takes the form of a reductio. (I'm going to explain his argument for the view first, and then explain precisely what it means to claim that the world is conceptually structured.) So, let's assume for now that McDowell's conclusion is wrong, the world is not conceptually structured. In Sellarsian talk, this would mean “that the space of reasons, the space of justifications or warrants, extends more widely than the conceptual sphere” (McDowell, 1994/1996, p. 7). Consider my belief *that there is a chair*. Intuitively we want the chair itself to play some role in warranting that belief, and so, as we've just seen, it must be within the “space of reasons”. If the chair (like the rest of the world) is not conceptually structured then that means the warrant for my belief *that there is a chair* goes beyond the conceptual sphere — beyond conceptually structured items —, and instead involves the non-conceptually structured object which is the chair. This is what McDowell means when he says that this view is committed to claiming that “the space of reasons . . . extends more widely than the conceptual sphere”. A natural view of this would be that what happens when I form the belief *that there is a chair* is that light bounces off the chair, into my eyes, causing my retina to send certain electrical signals to my brain, which eventually leads to my mind forming the conceptually structured belief *that there is a chair*. This picture holds that my relationship with the chair, up to the point that my mind gets involved and forms mental states with conceptual structure, is the result of a process involving the interaction of objects which are not themselves conceptually structured. But still, if my belief is to be justified by the world then it is the non-conceptual object — the chair — which provides the ultimate grounds for my belief in the chair's existence; the chair needs to be within the “space of reasons”, even though it is not within the conceptual sphere.

McDowell argues that this account is fatally flawed. His argument is based on the following claim:

we cannot understand the relations in virtue of which a judgement is warranted except as relations within the space of concepts: relations such as implication or probalification, which hold between potential exercises of conceptual capacities. The attempt to extend the scope of justification relations outside the conceptual sphere cannot do what it is supposed to do. (McDowell, 1994/1996, p. 7).

What's wrong with the account involving the non-conceptually structured chair is that there will be a step in the chain of justification — from the chair to my belief — that involves something which is not-conceptually structured supposedly justifying a conceptually structured item. McDowell claims that only contents or contentful items can stand in rational relations like justification to other contents or contentful items. If this is true then that means that there can be no justificatory chain which involves the non-conceptual object which is the chair at one end, and my belief *that there is a chair* at the other. And so, this view cannot account for how my belief *that there is a chair* can be justified by the chair in the world. The non-conceptual item which leads to my forming the belief *that there is a chair* is forced to look like “a brute impact from the exterior” (McDowell, 1994/1996, p. 8). And so, if items in the world lack conceptual structure, then it seems impossible for them to play a role in justifying our beliefs and we could never have justified beliefs about them.

McDowell's argument here depends upon the claim that justificatory relations can only hold between contentful (/conceptually structured) items,⁵ but why does he think this? The answer is found in McDowell support for epistemic internalism. According to epistemic internalism, for our beliefs to be justified we need to be able to self-consciously evaluate the warrant for our beliefs, or, as McDowell puts it, epistemic internalism is the view that:

the warrant by virtue of which a belief counts as knowledgeable is accessible to the knower; it is at least potentially known by her. Someone who has a bit of knowledge of the relevant sort is self-conscious about the credentials of her knowledge. As Sellars puts it, she occupies a position in the space of being able to justify what one says. (McDowell, 2011, p. 17).

In *Perception as a Capacity for Knowledge* McDowell (2011) argues for this view by contrasting it with Burge's epistemic externalism. According to Burge we don't need to have rational access to the warrant for our perceptual beliefs for our beliefs to be justified —

⁵McDowell's claim here — that justificatory relations can only hold between conceptually structured items — parallels Davidson's claim that “nothing can count as a reason for holding a belief expect another belief” (Davidson, 1983/2001, p. 141). The difference between Davidson and McDowell is that McDowell doesn't think that only beliefs can be conceptually structured, instead he thinks that facts and experiences can have conceptual structure, which explains how they are able to provide justification for beliefs. This enables him to avoid the central pitfall of Davidson's coherentism, which is the problem of explaining how our beliefs are rationally related to the outside world.

“The individual need not have the concepts necessary to think the propositional content that formulates the warrant” (Burge, 2003, p. 504). A classic, if overly simplified, epistemic externalist view would be something along the lines of the view that a belief is justified if it is a result of a causal process which typically leads to true beliefs. The perceiver doesn’t need to be able to have conceptual access to every element of this process for their beliefs to be justified by it, it just needs to be reliable. (I’m not going to go into the details of the argument between epistemic externalism and internalism, but it seems hard to deny that epistemic internalism would be the ideal position if we’re able to get it to work — we want to be able to understand and evaluate the justification for our beliefs — and McDowell provides a view of the world, and our relationship with it, which makes this possible. From now on I’m going to simply assume that epistemic internalism is true, and then look at McDowell’s argument about the consequences.)

So, if epistemic internalism is true, then our beliefs can only be justified by processes which we can rationally evaluate and conceptualise. This means that any causal process which involves an item which lacks conceptual structure and an item with conceptual structure (such as an experience or a belief) cannot be a justificatory relation because the non-conceptually structured item’s role will be left seeming to be “a brute impact from the exterior” (McDowell, 1994/1996, p. 8). For such processes there will be stages for which we lack the rational access that is required by epistemic internalism. And so, if epistemic internalism is true, rational relations like justification can only hold between contentful items. This means that if the world is not made up of contentful items then it can’t play a role in justifying our beliefs. We want our beliefs about the world to be justified by the items in the world themselves, so McDowell concludes that the items in the world must themselves have conceptual structure. And not only must the items in the world have conceptual structure, but every step in the causal process which leads to our forming our beliefs must also be contentful (so our experience of the world must also be conceptually structured).

It is the fact *that there is a chair* that causes me to have the experience *that there is a chair*, and this in turn leads to my believing *that there is a chair*. This explains how it is possible for “the world itself [to] exert a rational constraint on our thinking” (McDowell, 1994/1996, p. 42), since the world itself is made up of conceptually structured items, such as the fact *that there is a chair*. We are able to have (conceptually structured) experience of this fact, and that in turn can enable me to form a justified belief. The chain involves only conceptually structured items and so it is able to satisfy the demands of epistemic internalism because it is possible for us to self-consciously evaluate every item in the chain. According to McDowell if epistemic internalism is true then we must hold this view of the world, or be forced to accept that none of our beliefs about the world are ever justified.

3.3.1 The meaning of “conceptually structured”

So, what does McDowell mean by claiming that the world is conceptually structured? He describes the view as follows:

In a particular experience in which one is not misled, what one takes in is *that things are thus and so*. *That things are thus and so* is the content of the experience, and it can also be the content of a judgement: it becomes the content if the subject decides to take the experience at face value. So it is conceptual content. But *that things are thus and so* is also, if one is not misled, an aspect of the layout of the world: it is how things are. Thus the idea of conceptually structured operations of receptivity puts us in a position to speak of experience as openness to the layout of reality. Experience enables the layout of reality itself to exert a rational impact on what a subject thinks. (McDowell, 1994/1996, p. 26, emphasis in original).

But what does it mean to claim that “*that things are thus and so* is . . . an aspect of the layout of the world”? McDowell here is essentially advocating the view of reality which Wittgenstein endorses in the opening of the *Tractatus Logico-Philosophicus* (Wittgenstein, 1922) with the propositions:

I The world is everything that is the case.

I.I The world is the totality of facts, not of things. (Wittgenstein, 1922, p. 31).

Or, as McDowell himself puts it:

there is no ontological gap between the sort of thing one can mean, or generally the sort of thing one can think, and the sort of thing that can be the case (McDowell, 1994/1996, p. 27).

If the world is made up of facts and not things, that means that what ultimately justifies my belief *that there is a chair* is simply the fact *that there is a chair*. Since this fact is conceptually structured — it essentially involves concepts like *chair* — I am able to have a conscious awareness of that very fact (a conceptually structured experience), and of the justificatory relationship it has with my belief. This means that there can be a rational, justificatory, link between the fact itself and my belief, and so we can give an account of our relationship with the world which satisfies the demands of epistemic internalism.

By putting it in terms of the lack of an “ontological gap” between our thought and the world it sounds like McDowell could be making a radical metaphysical claim. Is he claiming that there are no objects like chairs in the world, only conceptually structured facts such as the fact *that there is a chair*? He cautions us against reading his view as anything other than truistic:

But to say there is no gap between thought, as such, and the world is just to dress up a truism in high-flown language. All the point comes to is that one

can think, for instance, *that spring has begun*, and that very same thing, *that spring has begun*, can be the case. (McDowell, 1994/1996, p. 27).

So, according to McDowell, all the *Tractarian* view of the world commits us to is the claim that the world is made up of the kinds of things we can talk and think about — facts. When we say ‘there is a chair’, according to McDowell’s view this is equivalent to saying ‘it is a fact that there is a chair’; by claiming that the world is made up of conceptually structured facts, McDowell does not mean to deny the existence of any ordinary objects, or claim that any of our common-sense claims about the world are wrong. He’s merely claiming that the ordinary objects which we talk about when talking about the world are apt for being talked about.

If the world is made up of facts, then we can explain how there can be a rational relationship between us and the facts in the world, and we can account for how they are able to justify our beliefs, even under an internalist account of justification. The view McDowell is supporting here is simply the view that:

When we say, and *mean*, that such-and-such is the case, we — and our meaning — do not stop anywhere short of the fact; but we mean: *this — is — so*. (Wittgenstein, 1953/2001, §95).

3.4 Conceptually structured inner awareness

McDowell’s solution to the rule-following problem essentially involves making the same argument about experience of the inner world as he does about experience of the outer world. He claims that:

That the inner world is a world at all consists in its conforming to the Kantian requirement: it is constitutively apt for conceptual representation, not something set over against a conceptual scheme. (McDowell, 1991/1998f, p. 310–311).

Our awareness of our own mental states, just like our experience of the external world, justifies many of our beliefs (such as the belief *that I mean plus by ‘+’*, or *that I’m intending to type a period*, or *that I’m currently in pain*). Just like experience of the outer world, it follows from epistemic internalism that if this awareness is to justify beliefs about our mental states then it must be conceptually structured.

So, the epistemological argument we saw in the previous section applies equally well to the case of our awareness of inner facts as it does to the outer world. This is incredibly useful because it enables McDowell to go beyond simply insisting that “there is a way of grasping a rule which is *not* an *interpretation*” (Wittgenstein, 1953/2001, §201, emphasis in original) and to flesh out a view of the inner which is not only motivated by the fact that

it enables a solution to the rule-following problem but also by independent epistemological arguments.⁶

Of course, McDowell is not claiming that the inner world of meanings, intentions and qualia is exactly like the outer world. He continues the passage quoted above by saying:

That it is inner consists in there being nothing to its states of affairs except the instantiation in consciousness of the relevant concepts; the instances of the concepts, unlike instances of concepts of the outer, have no being independently of that fact that the concepts they instantiate figure in the content of consciousness. . . . But that is not to say they have no being. (McDowell, 1991/1998f, p. 311).

I'm going to argue at the end of this chapter that this aspect of his view leads to serious problems for McDowell. But for now it's important simply to note that according to McDowell the difference between the inner and the outer worlds is that facts about the inner "have no being independently of that fact that the concepts they instantiate figure in the content of consciousness", whereas facts about the outer world are so independently of our experience of them — the fact that *earth is the third planet from the sun* would be so even if there were no conscious beings to be aware of it.

In the case of outer experience McDowell claims that when I see a chair my experience enables me to perceive the fact *that there is a chair*, and because this fact is conceptually structured it is able to be part of a causal process which justifies my belief *that there is chair*. There is no interpretation needed here, we can simply see the fact itself. This enables us to make sense of what McDowell means when he talks about our ability to "hear someone else's meaning in his words" (McDowell, 1984/1998o, p. 253). When we hear someone, let's call them John, saying '68 + 57 is 125' we are able to hear them saying *that 68 plus 57 is 125* — in my auditory experience I am experiencing the conceptually structured fact *that John is saying that 68 plus 57 is 125*. Because what I am experiencing is a conceptually structured fact the concept of *plus* is itself involved in that experience, and so that they have used the symbol '+' to mean *plus* (and not *quus*) can be part of the experience itself. This means that from the experience itself (the experience *that John is saying that 68 plus 57 is 125*) it is possible for me to know that they used '+' to mean *plus*. Another example which McDowell uses is that of sign-posts.⁷ Consider the case of looking at a sign-post pointing to the left, in that case our conceptually structured

⁶McDowell notes that he doesn't aim to be putting forward a view which is equivalent to the solution Wittgenstein himself advocates. Instead, he merely claims that it "makes good sense of some of what he says" but "it leaves some of what he says unexplained, and some looking positively mistaken" (McDowell, 1989/1998j, p. 280). See also (McDowell, 1991/1998f, p. 306, fn. 20 & p. 310, fn. 28).

⁷McDowell uses the example of sign-posts in "Wittgenstein on Following a Rule" (McDowell, 1984/1998o), "Intentionality and Interiority in Wittgenstein" (McDowell, 1991/1998f), and "How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein" (McDowell, 2002/2009d).

experience of the sign-post will be an experience of the fact *that the sign-post is pointing to the left* and from that experience itself — without interpretation — we are able to know which way the sign-post points. Crucially, in the case of inner awareness of our own meanings he makes precisely the same move. In the case where we are using a word (or remembering our previous use of a word), what we are aware of is the fact *that we meant plus by ‘+’*. Our awareness involves the concept of *plus* itself, and so there is no way for Kripke’s sceptic’s doubts to get off the ground.

As we saw in §3.2, in rejecting the master thesis McDowell aims to make space for the idea that things with normative significance can come to mind. The claim that our inner awareness involves conceptually structured items enables this because the concepts that are involved in inner awareness of our meanings, such as *plus*, have normative significance. Concepts like *plus* themselves have the required ability to sort things into those which fit with them and those that do not; it is inherent to the concept of *plus* that *68 plus 57 is 125* is correct, and *68 plus 57 is 5* is not. If our awareness of the meanings we currently associate with words (and our memory of our past meanings) is conceptually structured and involves concepts like *plus* then we can simply refer to the fact *that I meant plus by ‘+’* when responding to Kripke’s sceptic, since we can have experience of fact itself.

3.5 Knowledge of meaning and epistemic fallibilism

One problem with McDowell’s solution here is that it seems to make the response to Kripke’s sceptic too easy. We sometimes make mistakes about what someone means when we hear them speak, and McDowell is guilty of simply ignoring this possibility. Consider a speaker who hadn’t correctly grasped the meaning of ‘+’ and actually used ‘+’ to mean *quus*. If we came across such a speaker and the first thing we heard them say was the sentence ‘I wonder what $68 + 57$ is?’ we would form the belief that what they were saying was *I wonder what 68 plus 57 is*. But that would be wrong, since what they actually were saying was *I wonder what 68 quus 57 is*. How can McDowell account for the fact that we will make mistakes like this? If he can’t provide an account of this possibility then his position seems guilty of simply assuming that we will always hear people to be associating the correct meanings with their words, which is clearly false.

McDowell himself considers a possibility just like this in the paper “Anti-realism and the Epistemology of Understanding” (McDowell, 1981/1998b):

Suppose Martians look exactly like human beings, and every sentence of their language sounds exactly like an English sentence, but, in general, means something quite different. A short burst of Martian speech might chance to occur in such circumstances that it seemed to be intelligible as English . . . Now if Dummett’s picture of the epistemology of understanding were correct, it ought to be a serious question how it can be that, say in a brief conversation with

a total stranger, one really understands his utterances. For one would need to be able to exclude, purely in terms of what is available to one at the putatively foundational level, the possibility that he is just such a Martian; and has one enough, at that level, to exclude that possibility? Once one has this worry, extending to any finite length the periods during which things seem to go smoothly cannot make any difference of principle. So it comes to seem that any claim to understand someone's utterances goes beyond what we have any genuine epistemic right to. (McDowell, 1981/1998b, p. 337).

In this case the Martians, like the quusser, use words which sound just like our own but have different meanings. This means that when hearing a short burst of their speech we can mistake it for our own language and think that we understand it. For example, consider the case where we come across a Martian saying a sentence which sounds like 'Snow is white'. McDowell points out that since they look and sound just like a normal human English speaker we are likely to then form the belief that we can hear them saying *that snow is white*. But in actual fact these words have completely different meanings in Martian and what they were actually saying was *that the table is large*. If we are susceptible to making mistakes like this then what right do we have to say that when hearing an English speaker we can simply "hear someone else's meaning in his words" (McDowell, 1984/1998o, p. 253)? If, for all I know, an English speaker could be speaking Martian, then how can I ever know that someone is saying *that snow is white*?

McDowell responds to this worry by arguing that it is based on endorsing a deeply flawed epistemological principle. This is the view that knowledge is only possible in cases where our evidence provides us with a *guarantee* that makes error impossible. He responds as follows:

if that idea is an application of a general epistemological principle, it must be one on these lines: the title of a state to count as knowledge depends on there being facts, unproblematically available to its possessor, that constitute a *guarantee* that the content of the putative knowledge is true. But as a general requirement for knowledge, this is worthless. Consider, for instance, the acquisition of knowledge by hearsay. Generally speaking, if one's informant knows what he tells one, and one understands what he says, one acquires his knowledge at second hand; whereas if he does not know, what one acquires cannot be knowledge. It does not undermine the possibility of acquiring knowledge by testimony that no informant is absolutely reliable (McDowell, 1981/1998b, p. 335, emphasis in original).

Here he argues by pointing out that it is possible to know things by hearing them though testimony, even though testimony can never provide us with a guarantee that the belief we have formed is not incorrect. If we accept that we can get knowledge from testimony then we must reject the epistemological principle which claims that knowledge requires evidence which *guarantees* the relevant belief to be true, and instead adopt a form of epistemological fallibilism.

In response to the problem of the possibility that a supposed English speaker may actually be speaking Martian, McDowell points out that if knowledge did require a guarantee that we are not talking to a Martian then the understanding of other's speech would always be impossible, since we could never absolutely guarantee that another speaker is speaking our own language. But if we instead accept a form of epistemic fallibilism then in most ordinary cases we will have a right to take it for granted that people who sound like us are speaking English, and in cases where this is indeed the case we are able to know what they are saying simply by hearing their speech. As he says in "Wittgenstein on Following a Rule":

My right to claim to understand him is precarious, in that nothing but a tissue of contingencies stands in the way of my losing it. But to envisage its loss is not necessarily to envisage its turning out that I never had the right at all. (McDowell, 1984/1998o, p. 252).

We saw an argument for the same epistemological claim in chapter 1. There the problem was how we could know that the object before us was a tomato, just by looking, since we are unable to rule out that it is not in fact a cunning façade. McDowell rejected the epistemological picture which claimed that knowledge requires that we rule out any possibility that we were making a mistake, and instead claimed that "experience is an essentially fallible, but unmediated, openness on our part to a reality that includes substantial (non-flat) states of affairs" (McDowell, 1989/1998g, p. 362). When all goes well, such as when we look at an actual tomato in normal conditions, our experience of the tomato enables knowledge of the fact that *that is a tomato*. The fact that this ability is fallible and sometimes leads to us forming false beliefs (since things which aren't tomatoes can look like tomatoes) doesn't mean that we can't acquire knowledge by looking in cases where all goes well.

This solution enables us to justify McDowell's claim that we are able to "hear someone else's meaning in his words" (McDowell, 1984/1998o, p. 253). We have a fallible capacity to do this, and in the good case where we are in normal conditions and the speaker is an English speaker with a normal understanding of the words they are using we can simply hear in experience of their talk what they say and mean. Whilst it is true that we will sometimes make mistakes, that does not mean that we can never directly hear what someone means.

3.5.1 Interpretations and disjunctivism

An interesting aspect of the role McDowell's epistemology plays in his response to the rule-following problem is the parallel between disjunctivism in his epistemological picture and his endorsement of the claim that "there is a way of grasping a rule which is *not* an *interpretation*" (Wittgenstein, 1953/2001, §201, emphasis in original).

The epistemic view which McDowell's opposes is sometimes referred to as the "highest common factor" view (McDowell, 1982/1998d, p. 388). This is the view that if our experience would *seem* the same to us in two different situations, such as when looking at a tomato or a tomato façade, then our experience itself can't involve any more than what would be available to us in the bad situation. And because seeing the tomato and the façade are subjectively indistinguishable then even in the good case the experience can't involve direct awareness of the fact *that there is a tomato*.

McDowell's direct realist view of experience rejects this and instead holds that when all goes well in our experience "the fact itself is directly presented to view" (McDowell, 1982/1998d, p. 387). So, when looking at a tomato, if we are in normal light, etc., and have no reason to doubt that what is in front of us is actually a tomato then our experience itself is an experience *that there is a tomato* — the fact *that there is a tomato* is an aspect of our experience itself. And he argues that this is so even though we may sometimes make a mistake and think that we can see *that there is a tomato* when what we are actually looking at is a mere façade.

The epistemological views which McDowell rejects hold that all that our experience alone presents us with is the fact that we can see something which looks like a tomato. The experience itself can't involve the fact that *there is a tomato* because we don't know for certain that we're not actually seeing a façade. According to this view, to have knowledge that there's a tomato we need to go further and interpret the limited experiential data to infer that there is a tomato in front of us. This is interesting because this epistemological view seems to directly parallel the view which McDowell opposes in the case of meaning — the view that in order to know what someone means when we hear them speak (or to know what we meant in the past, or . . .) we need to *interpret* our inconclusive experiential data (or memory, or . . .). But, as we saw above, if such interpretation is necessary at any stage then whatever comes to mind will need further interpretation in order to have a meaning. And so we get an infinite regress. In response McDowell rejects the master thesis and endorses Wittgenstein's claim that "there is a way of grasping a rule which is *not an interpretation*" (Wittgenstein, 1953/2001, §201, emphasis in original). Instead he holds that we can *directly* hear someone's meaning in their words, or remember what we meant in the past. We are able to hear someone saying precisely *that 68 plus 57 is 125*, even though it may sound just like a Martian saying something completely different. This directly parallels his direct realist account of experience — that when looking at a tomato we are able to see that *that is a tomato*, even though the experience is subjectively indistinguishable from looking at a tomato façade.

3.6 Quietism

As we saw in section §3.1 (“McDowell’s diagnosis of Wright’s mistake”) Wright accuses Wittgenstein of a philosophically unsatisfying quietism. Wright reads Wittgenstein as completely rejecting the idea of inner awareness, hiding behind “quietism” as an excuse to avoid giving us anything to replace it. McDowell disagrees with this reading of Wittgenstein and instead argues that Wittgenstein only rejects the *Cartesian* account of inner awareness. Instead he claims that Wittgenstein endorses an alternative view of inner awareness, which he claims to be common sense. And according to McDowell once we’ve endorsed this commonsensical view of the inner then our ability to know what we mean (and meant in the past) is left looking completely unmysterious and prevents the question of how meaning is possible from being troubling at all. According to McDowell, Wittgenstein’s general philosophical strategy⁸ is as follows:

Wittgenstein does not amass philosophical doctrine, or point to areas where it would be a good thing for others to do that, though “quietism”⁹ debars him from doing so. He uncovers tendencies to forgetfulness, which lead to trouble when we engage in philosophical reflection, but which can be counteracted by suitable reminders of the obvious. (McDowell, 2002/2009d, pp. 105–106).

In the specific case of the apparent problems involving meaning McDowell claims that:

The right response to “How is meaning possible” . . . is to uncover the way of thinking that makes it seem difficult to accommodate meaning and intentionality in our picture of how things are, and to lay bare how uncompulsory it is to think in that way. (McDowell, 1993/1998h, p. 272).

But if all Wittgenstein is doing is reminding us of a piece of common sense, what about all McDowell’s talk of conceptual structure? McDowell argues that experience of the outer world must be conceptually structured, and that if our awareness of the inner is also conceptually structured then we can solve the rule-following problem. This seems to

⁸It’s important to note that McDowell does not think that all philosophy should be quietist. He does think that many apparent philosophical problems are a result of confusion which once removed makes the problems no longer pressing. But he also thinks that constructive philosophy is sometimes appropriate, such in the case of “reflection about the requirements of justice or the proper shape for a political community” (McDowell, 2009e, p. 367).

⁹In this particular paper McDowell often uses the word “quietism” to talk about quietism in a pejorative sense, the kind of philosophical laziness of which Wright accuses Wittgenstein. But elsewhere McDowell uses “quietism” in a positive sense, to refer to the aspects of Wittgenstein’s strategy which he endorses — the strategy of reminding us of common sense to prevent philosophical problems from appearing pressing. It is in this second sense that I’ll use “quietism” in this chapter. (For an example of McDowell using “quietism” in this positive sense, see “Wittgensteinian “Quietism” ” (McDowell, 2009e).

go far beyond merely “[uncovering] tendencies to forgetfulness” (McDowell, 2002/2009d, p. 105). In order to explain why McDowell claims that his use of conceptually structured inner awareness fits with Wittgenstein’s philosophical strategy, which “leaves everything as it is” (Wittgenstein, 1953/2001, §124), it’ll be useful to look at what McDowell is *not* doing.

McDowell contrasts his (quietist, Wittgensteinian) response to the rule-following problem with a philosophically constructive (and so non-quietist, and non-Wittgensteinian) response which would involve “[concocting] substantial philosophical answers to [the questions] “How is meaning possible?” or “How is intentionality possible?” ” (McDowell, 1993/1998h, p. 272). Instead he argues that:

Wittgenstein’s aim is not to substitute *something else*, the community’s conferring of authority to individuals, for the idea of a special relation to the subject-matter of their remarks, as that which makes them specially authoritative in making them. Wittgenstein’s aim is, rather, to disarm the temptation to think that the latter idea requires *queer* items or states of affairs. (McDowell, 1989/1998j, p. 291, emphasis in original).

What’s key to McDowell’s response is that, unlike philosophically constructive responses, he’s “not giving an account of norm-governedness as such” (McDowell, 2002/2009d, p. 104) — he’s not giving us an account of what constitutes meaning. And in evaluating whether or not McDowell sticks to Wittgenstein’s quietist methodology it’s essential to remember that he does not do this. McDowell does not attempt to *explain* how conceptually structured experience of the outer world is possible, he just argues that it must be if we are going to have knowledge of the world at all. And similarly, he doesn’t explain how conceptually structured awareness of our inner states is possible, instead he attempts to undermine philosophical views which imply that it can’t be possible, and argues that it must be possible if we are able to have justified knowledge of our inner states and if we are going to be able to avoid the problem of regress of interpretations threatened by Kripke’s sceptic. It is in this lack of answers to such “how possible” questions that McDowell’s philosophical methodology is true to Wittgenstein’s quietism.

In contrast, remember Wright’s response to the rule-following problem. Wright attempts to give an account of the facts which are *constitutive* of having a particular meaning in mind — best judgements are what *make it the case* that you, say, mean *plus* by ‘+’. This is a direct answer to the question of how it is possible to have a particular meaning in mind, because it gives an account of what makes it true that you, say, mean *plus* by ‘+’. This is precisely what McDowell does not do.

The common sense about the inner world which McDowell is trying to restore is the claim that:

human beings can be initiated into the capacity to place themselves within a “normatively” structured space of possibilities (McDowell, 1991/1998f, p. 319).

As we saw towards the end of §3.2, McDowell does this by arguing that the master thesis (“the thesis that whatever a person has in her mind, it is only by virtue of being interpreted” (McDowell, 1993/1998h, p. 270)) is highly counterintuitive. He points out that the master thesis implies that intentions and meanings themselves cannot come to mind, or, at least, if they can come to mind it is only through interpretation that they are able to have meaning or normative force. He argues that once we’ve realised how contrary to common sense this is, and instead recognise that meanings and intentions can come to mind, then the question of “How is meaning possible?” becomes nonthreatening — Question: “How is it possible to know you meant *plus* by ‘+’ yesterday?”; Answer: “I can remember the concept *plus* coming to mind when using the ‘+’ symbol”. Of course this answer doesn’t actually tell us *how meaning is possible*, but if we accept that meanings can come to mind and be publicly communicated (the piece of common sense which some philosophers are apt to forget), then the lack of an answer to that question will at least no longer be pressing, and Kripke’s sceptic will not seem to threaten the very idea of meaning.

We can give a similar account of McDowell’s claim that our experience of the outer world is conceptually structured. As I noted above, McDowell’s claim that “there is no ontological gap between the sort of thing one can mean, or generally the sort of thing one can think, and the sort of thing that can be the case” (McDowell, 1994/1996, p. 27) is not meant as a radical metaphysical claim. Instead, it is nothing more than the truism that:

one can think, for instance, *that spring has begun*, and that very same thing, *that spring has begun*, can be the case. (McDowell, 1994/1996, p. 27).

And again, he intends this to be simple common sense. And if we accept the truth of this piece of “common sense” then the epistemological problems he was concerned with simply fade away.

So — at least if we accept McDowell’s claim that it is simply common sense to acknowledge that we are able to bring intentions to mind and that the world is made up of facts — it seems clear how his position is quietist. Once he has reminded us of these pieces of common sense we should no longer feel threatened by questions like “How is knowledge of the external world possible?” or “How is meaning possible?” because we will no longer be worried that such things might not be possible.

However, shouldn’t we still be interested in how these things are possible? Even if we believe it is common sense that conceptually structured items can come to mind and that we can simply hear meanings in other people’s talk, wouldn’t we still be intrigued how this is possible? Consider the hard problem of consciousness, even if you’re convinced

that we are conscious and that we can have experiences with phenomenal character, that doesn't mean that the question "How is consciousness possible?" isn't one that's worthwhile investigating.

Towards the end the book *An Introduction to Contemporary Metaethics* Miller (2003a) makes a point along similar lines:

curiosity should be regarded as healthy in the absence of a positive reason to think otherwise: in the absence of such a reason, the curiosity stands as the default position. McDowell's assumption that a piece of curiosity has no standing unless we have a demonstration that is healthy gets it the wrong way round. (Miller, 2003a, p. 282).

In contrast, McDowell seems to think that we should only be interested in "How is x possible?" questions if we have reason to believe that x is not possible:

If a question of [the form 'How is it possible that ...?'] is to express a determinate philosophical difficulty, it must be asked from a frame of mind in which there is at least a risk of its looking as though whatever the question is asked about is *not* possible. So one's first move, if someone tries to interest one in a 'How is it possible?' question, should be to ask; why exactly does it look to you, and why should it look to me, as if such-and-such a thing (e.g. a baseless authority about oneself) is *not* possible? If the response to this follow-up question discloses a frame of mind that can intelligibly and interestingly captivate one, then this question poses a worthwhile task for philosophy. For contrast, consider a crude version of a kind of problem that bothered some ancient Greeks (though I need not claim historical accuracy here). The question is: 'How are false statements possible?' What makes it look as if false statements might not be possible is a conception according to which meaningfulness in general, including the meaningfulness of statements, is a matter of a relation — something along the lines of the name-bearer relation — to something actual. Such a difficulty in making room for falsehood does not (at least, not any longer) pose a worthwhile task for philosophy. (McDowell, 2000, pp. 57–58, emphasis in original).

At first sight, McDowell's use of the example of the ancient Greeks' worries over the possibility of false statements here seems problematic. In response to someone in the grip of this worry we could do much more than dispel them of their confusion about meaning which made it look like that only statements which are true are possible. We could give them an account of a theory of meaning which made it clear how statements in general are meaningful, and how this is much the same for both true and false statements. There are many alternative theories of meaning on offer, and fleshing them out seems to be the kind of constitutive philosophical task which McDowell's quietism would oppose. It's not at all clear why recognition of the commonsense facts which make it clear that such statements are possible shouldn't prevent us looking into how statements are possible. However, this response is far too quick, and much more would be required for a thorough response to McDowell's quietism, but doing so is beyond the scope of this thesis.

3.7 McDowell's solution as involving *sui generis* mental states

We saw in §2.1.2 that after rejecting the ideas that qualia or mental images could be the facts that constitute meaning, Kripke goes on to briefly consider the idea that the facts which constitute meaning could be some wholly different kind of mental state. This is the idea that:

Perhaps [meaning *plus* by ‘+’] is simply a primitive state, not to be assimilated to sensations or headaches or any ‘qualitative’ states, nor to be assimilated to dispositions, but a state of a unique kind of its own. (Kripke, 1982, p. 51).

McDowell's response to the rule-following problem could be seen as taking a move precisely along these lines — meanings and intentions are mental states which are neither dispositions or qualitative states like headaches.

Kripke's criticism of this view starts by attempting to show that this *sui generis* mental state can't be introspectible because “if there really were an introspectible state ... it would have stared one in the face and would have robbed the sceptic's challenge of any appeal” (Kripke, 1982, p. 51). But he accuses the idea of an non-introspectible state, wholly different from all other mental states — “a unique kind of its own” — as seeming “desperate” because “it leaves the nature of this postulated primitive state — the primitive state of ‘meaning addition by “plus” ’ — completely mysterious” (Kripke, 1982, p. 51).

But surely knowledge of our meanings does stare us in the face! It is only Kripke's reductionist assumption which blocks us from referring to knowledge of our meanings to respond the sceptical challenge; once that's been removed can't we simply refer to our knowledge of the fact that *I mean plus by ‘+’*? Also, as McDowell points out when attacking the master thesis, there is nothing strange about the idea of a mental state which has normative properties (i.e. tells us how we should act, such as how to use ‘+’ in new situations), since if there are no such mental states then not only could meanings not come to mind but neither could intentions or rules.

Kripke's second criticism of this view is the argument that the mental state would have to be finite, because it is contained in our finite minds, and so how could it determine the correct usage of ‘+’ for *all* numbers? And he claims that there is no reasonable way to “conceive of a finite state which *could* not be interpreted in a quus-like way” (Kripke, 1982, p. 52, emphasis in original).

But, as we've already seen, McDowell's solution to the rule-following problem depends upon claiming that there is a way of following a rule which is not an interpretation. The *sui generis* mental state involved must not require interpretation to have meaning, since then, as Kripke points out, there would always be the possibility of interpreting it in a

quus-like way. But if we reject the master thesis and accept that things can come to mind which do not require interpretation to have a meaning then this is not a problem.

But what about the worry that this involves having something infinite come to mind? In response to this worry it's important to remember the distinction between concepts themselves and their extension. Grasping the concept of *plus* does enable us to know how to add any combination of numbers, but that doesn't mean that we need to have the correct answer to all addition problems come to mind when we think of *plus*. If you correctly understand *plus* then you will be able to know, in any particular case,¹⁰ what answer you should give; but there's no reason to believe that every single one of these cases must come to mind whenever you think of *plus*. In rejecting the master thesis McDowell endorses the idea that things can come to mind which, without interpretation, determine what we should do in different cases. Once we've recognised that this doesn't involve having the entire extension come to mind whenever we consider a concept, there seems to be no reason to believe that finite beings such as ourselves cannot do this.

3.8 The role of the community

One problematic aspect in interpreting McDowell's solution to the rule-following problem is to understand the role the community plays. In his early paper on rule-following, "Wittgenstein on Following a Rule", he says:

The diagnosis prompts the question "How can there be a way of grasping a rule which is not an interpretation?", and I think the thesis that obeying a rule is a practice is meant to constitute the answer to this question. (McDowell, 1984/1998o, p. 238).

He continues:

How can a performance both be nothing but a "blind" reaction to a situation, not an attempt to act on an interpretation ... and be a case of going by a rule ...? The answer is: by belonging to a custom [(Wittgenstein, 1953/2001, §198)], practice [(Wittgenstein, 1953/2001, §202)], or institution [(Wittgenstein, 1956/1978, VI-31)]. (McDowell, 1984/1998o, p. 242).

This solution to the rule-following problem seems to make essential use of our membership in a linguistic community:

Wittgenstein's problem was to explain how understanding can be other than interpretation ... This non-anti-realist conception of a linguistic community gives us a genuine right to the following answer: shared command of a language equips us to know one another's meaning without needing to arrive at

¹⁰Or at least you will be able to know the answer to any addition problem which is simple enough for you to be capable of working it out in your lifetime.

that knowledge by interpretation, because it equips us to hear someone else's meaning in his words. (McDowell, 1984/1998o, p. 253).

It seems clear here that in this paper McDowell is claiming that the fact that we're members of a community with a shared command of a language is essential to his response to the rule-following problem. Such a response would be problematic in fitting in with McDowell's overall view of how we should approach the problem of meaning because it would constitute a constitutive (i.e. non-quietist) response to the question "How is meaning possible?" This is particularly obvious in the first of these three quotations, since if he thinks that we should answer this question by talking about the way "obeying a rule is a practice" then it looks like he is actually going towards answering the question, rather than merely showing why the worries behind the question depend upon forgetting common sense. In the much later paper "Meaning and Intentionality in Wittgenstein's Later Philosophy" McDowell (1993/1998h) himself criticises what he said in "Wittgenstein on Following a Rule" for appearing to be guilty of this very mistake.¹¹

Commentators often suggest that the concept of *custom* and its cognates figure in Wittgenstein as elements in a constitutive philosophical response to questions like "How is meaning possible?" According to some versions of this reading, Wittgenstein actually gives the response; according to others, he points towards it but does not give it, out of a quietism that must stand exposed as inappropriate by the sheer fact that the questions are (supposedly) good ones. I am committed to regarding this as a misreading. [Footnote:] I now think that ["Wittgenstein on Following a Rule"] is too hospitable to this kind of reading. (McDowell, 1993/1998h, p. 275, emphasis in original).

But at this point he doesn't reject all talk of the community in explaining his response to the rule-following problem, for example in the even later paper "How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein" he says that:

to be capable of being told what to do by a sign-post one needs to have been initiated into an appropriate practice (McDowell, 2002/2009d, p. 101).

So, it's clear that, at least after he wrote "Meaning and Intentionality in Wittgenstein's Later Philosophy", McDowell does not endorse using the community to provide a constitutive answer to the question "How is meaning possible?" But, even after writing that paper, he continues to talk a lot about the community when talking about meaning. So, what role does the community play in McDowell's view, if it is not to provide a constitutive answer to the question of how meaning is possible?

¹¹Indeed, in "Meaning and Intentionality in Wittgenstein's Later Philosophy" McDowell provides an argument to show that the community *can't* provide a constitutive response to the rule-following problem. This is because if we used custom to give an account of how meaning is possible then the customs used in that account "would need to be characterizable in terms that do not presuppose meaning and understanding" (McDowell, 1993/1998h, p. 276), which, of course, we cannot do.

A crucial clue here is McDowell's reference to Wittgenstein's claim that "You learned the *concept* pain when you learnt language" (Wittgenstein, 1953/2001, §384) in "One Strand in the Private Language Argument" (McDowell, 1989/1998j, p. 293). Here he's endorsing the claim that the concept *pain* (like all concepts) is acquired though learning language.¹² In "How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein" McDowell makes the same point about our ability to understand sign-posts: "not everyone who encounters a sign-post gets told which way to go" (McDowell, 2002/2009d, p. 101), and notes that Martians may use sign-posts to point in the opposite direction. The central point here is that our ability to understand which way sign-posts point (like our grasp of the concepts of *pain*, *plus* etc.) is learnt when learning language. If we had never learnt a language then we'd never have acquired these concepts. So although we can't use the community to provide us with an *explanation* of how meaning itself is possible, the existence of our linguistic community was needed for us to learn any particular meaning (or how to interpret sign-posts, etc.).

However, this leaves open the question of whether it would be possible for someone to acquire a concept without a linguistic community. Could a person born in isolation develop their own language? Just because it's the case that our concept of *plus* was acquired by being taught English, that alone doesn't show that it would be impossible for someone to grasp meanings in isolation.

3.9 The seems right / is right distinction

We saw in the last chapter that one key component of a successful account of meaning is being able to make sense of the possibility of making mistakes. It is essential that we can draw a distinction between what seems right to the speaker (or community) and what actually is right. In this section I'm going to distinguish between two types of mistake which a view needs to be able to make sense of: mistakes concerning how to properly apply a word, and mistakes in the speaker's judgement about which meaning they have in mind. I shall argue that although McDowell can easily make sense of the first, the second is problematic because his views on the nature of inner states seem incompatible with the possibility of making such mistakes.

As we saw in §3.4 ("Conceptually structured inner awareness"), according to McDowell a key difference between the inner and the outer worlds is that facts about inner states "have no being independently of that fact that the concepts they instantiate figure in the content of consciousness" (McDowell, 1991/1998f, p. 311). So, how does McDowell think

¹²This view of pain has led to worries about how to make sense of the possibility of the pain of pre-linguistic infants and non-linguistic animals. For McDowell's response to this worry see "One Strand in the Private Language Argument" (McDowell, 1989/1998j, §7) and *Mind and World* (McDowell, 1994/1996, pp. 119–123).

it is possible to draw the seems right / is right distinction for inner states? His answer is as follows:

That it is inner consists in there being nothing to its states of affairs except the instantiation in consciousness of the relevant concepts . . . But that is not to say that these states of affairs have no being. It is true that no distinction between “seems right” and “is right” opens up, with respect to the obtaining of these states of affairs, from the subject’s point of view. But what it is for the relevant concepts — those whose instantiation in consciousness exhausts the being of these states of affairs — to have application cannot be understood exclusively from the subject’s point of view. . . . So the distinction between “seems right” and “is right” opens up, with respect to these states of affairs, from the necessarily thinkable second-person or third-person point of view; one’s inner world is part of the world (anyone’s world). (McDowell, 1991/1998f, pp. 311).

McDowell here accepts that from the subject’s own *first-person* point of view there is no scope for them to draw the seems right / is right distinction. That seems reasonable — if it seems to me that in answering ‘5’ to ‘What’s $68 + 57$?’ I’m giving the correct answer then there’s no way I’ll be able to tell from my own point of view, at least at that moment, that my answer was actually wrong. I could never know at the same point in time that what is right is different from what seems right to me at that moment. In order to save our ability to make sense of the seems right / is right distinction McDowell brings in second and third-person points of view. Even though I can’t tell myself, at the time, that this response is incorrect that doesn’t prevent us from making sense of the idea that I could be making a mistake, because others can be in a position to judge that I have given the wrong answer — “the distinction between “seems right” and “is right” opens up, with respect to these states of affairs, from the necessarily thinkable second-person or third-person point of view” (McDowell, 1991/1998f, pp. 311). This leads McDowell to conclude that such concepts are “necessarily thinkable [from the] second-person or third-person point of view” (McDowell, 1991/1998f, pp. 311), in order for there to always be scope to draw the required seems right / is right distinction. So, in the case of mistakes regarding how to properly apply a word we can open up the seems right / is right distinction by pointing out that other speakers can recognise the fact that I’ve used the word incorrectly.¹³ If McDowell depends upon the perspective of other speakers in

¹³In a footnote shortly after the above quotation McDowell adds:

I do not mean this formulation to imply that the points of view must be occupied by different thinkers. I myself must be able to think of my being thus and so as a case of someone’s being thus and so. The actual existence of others is not what matters (at any rate not at this point in the overall picture). (McDowell, 1991/1998f, p. 312, fn. 31).

Here he claims that we can get these alternative points of view without requiring other people to exist. He doesn’t elaborate on this point here, but in *Mind & World* when discussing this issue he notes that these second or third person perspectives could be occupied by the person “herself at different times”

our community to make sense of the seems right / is right distinction, then what about the possibility of the whole community going wrong, how can he account for that? This isn't a problem for McDowell because, unlike Wright's communitarian view from 1980, the facts about meaning are not fixed by the community's dispositions. The community are involved here simply to show how we are able to *make sense* of the idea of a speaker making a mistake, and we can equally well make sense of the possibility of a whole community making a mistake. Let's say the community regularly referred to horsey-looking cows on dark nights as 'cows'. On McDowell's view the facts about the community's dispositions do not determine the extension of 'cow', instead what determines the extension of cow is simply the concept of *cow*. He can demonstrate that we can make sense of the possibility of the entire community using 'cow' incorrectly simply by pointing out that we are able to recognise that our past usage of a word was wrong. And even if in a particular case we haven't actually recognised that we've made a mistake, that is not significant since all he needs is the claim that we can *make sense* of such a possibility, and we can clearly do that. By requiring that all concepts are "necessarily thinkable [from the] second-person or third-person point of view" (McDowell, 1991/1998f, pp. 311) he secures the fact that it is always possible that there is another perspective on our use of a concept, from which it may become clear that what *seemed* to be correct use is not actually correct.

But there is another type of case where we want to draw the seems right / is right distinction, that of where a subject is judging not how to *apply* a particular concept, but *which* concept they have in mind. Consider the case where a subject judges *that they have formed the intention to give up smoking*;¹⁴ how does the seems right / is right distinction open up there? At first sight it may seem that we can simply make the same move as we did in the previous case, we can make sense of the possibility of mistakes because a subject may well judge from a first person point of view that they've formed this intention, but from second or third-person points of view it can become clear that in fact they've not formed this intention at all. But this response is problematic because it conflicts with what McDowell says about inner states elsewhere. Consider the following passages:

(McDowell, 1994/1996, p. 38). So, presumably the idea here is that it's possible for a subject to realise that their past use of a word was wrong. In this case they would have a second/third-person perspective on their own past usage of the word, and so are now able to realise that what seemed right to them at the earlier time was in fact wrong. So it seems that, according to McDowell, we can make sense of the seems right / is right distinction without bringing the community in at all.

¹⁴I use the example of the intention to give up smoking because it's a clear example of a case where someone may be mistaken about what intention they have formed. But what I say here should apply equally to cases involving meaning — a speaker may believe themselves to mean *plus* by '+', but actually mean *quus*.

That it is inner consists in there being nothing to its states of affairs except the instantiation in consciousness of the relevant concepts; the instances of the concepts, unlike instances of concepts of the outer, *have no being independently of that fact that the concepts they instantiate figure in the content of consciousness*. . . . But that is not to say they have no being. (McDowell, 1991/1998f, p. 311, emphasis added).

the objects of “inner sense” are internal accusatives to the awareness that “inner experiences” constitute; they *have no existence independently of that awareness* (McDowell, 1994/1996, p. 21, emphasis added).

If we can make out that the *judgements of “inner sense”* are about anything, it has to be that they *are about the impressions of “inner sense” themselves*, not about something independent of which the impressions constitute awareness. (McDowell, 1994/1996, p. 22, emphasis added).

If there is nothing to it being the case that someone intends to give up smoking other than the fact that the intention to give up smoking figures in their consciousness, then how can their conscious inner experience of intending to give up smoking possibly be wrong? On this view it seems that *all there is* to the facts about what someone intends is what figures in their current first-person conscious experience — what makes it the case that I am having a particular intention is the fact that I am experiencing having that intention. And so it seems that such experience can never be misleading, and there’s no scope for the seems right / is right distinction to open up here.

It seems that McDowell is committed to endorsing what Smith calls the “strong phenomenological conception” (Smith, 1994, p. 105) of inner items.¹⁵ According to this view a subject desires (or intends, or feels, or . . .) to ϕ if and only if she believes that she desires (or intends, or feels, or . . .) to ϕ .¹⁶ If this is the case then the seems right / is

¹⁵In this passage Smith is in fact only talking about the “strong phenomenological conception of *desires*” (Smith, 1994, p. 105, emphasis added), but the account seems appropriate for McDowell’s view of the inner in general.

¹⁶At this point I should note that Smith claims that McDowell rejects the phenomenological conception of desires, but he is cautious about this claim and notes that “[McDowell] nowhere says that he rejects such a conception” (Smith, 1994, p. 207, fn. 5). Smith argues for reading McDowell as rejecting the phenomenological conception because in the paper “Are moral requirements hypothetical imperatives?” McDowell asserts that “consequentially ascribed desires are indeed desires” (McDowell, 1978/1998c, p. 89). Smith argues that “the point of consequentially ascribed desires is that there may be no phenomenological ground for their ascription” (Smith, 1994, pp. 207–208, fn. 5), and so concludes that McDowell must reject the phenomenological conception of desires. At this point it seems unclear how we should read McDowell; perhaps he changed his mind about the objects of inner sense? He did write the paper Smith refers to back in 1978, whereas the three passages I quote above which indicate his support for the phenomenological conception are all from work published in the 90s. Or perhaps, it is significant that McDowell never explicitly rejects the phenomenological conception of desires, and Smith is wrong to read him as rejecting it in the first place. I’m going to simply ignore what McDowell says in the 1978 paper, and focus instead on the natural reading of the passages in the papers in which McDowell talks about rule-following.

right distinction will not be able to open up, since what is right is determined by what seems right.

One way to partially solve this problem is to point out that McDowell does not claim that *all* inner experiences constitute some inner fact. So he may hold that all inner facts are constituted by some inner experience, but not be committed to the view that all inner experiences constitute some inner fact. For example, let's say that I'm deceiving myself into believing I've formed the intention to give up smoking. In this case I could seem to have the experience of forming the intention to give up smoking, but I've tricked myself, and in that case I haven't actually formed any new intention regarding my smoking habits at all. So, although I have the experience of forming an intention, not all inner experiences constitute the fact they appear to, and so it's perfectly possible that I haven't actually formed that intention after all. Were I to have actually formed that intention I may have an inner experience which seems just the same to me at the time, but in the good case I wasn't deceiving myself, and so that experience *does* constitute the fact that I have formed the intention to give up smoking. This view has strong parallels to Wright's judgement-dependent account, in that experiences, like judgements in Wright's account, can constitute the inner facts, but not all experience or judgements will necessarily constitute such facts.¹⁷

This means that we can open up a distinction between it seeming to a subject that they intend to ϕ and actually intending to ϕ , because when experiencing an intention it is possible to falsely think that they have formed an intention when they have not.

However, despite its ability to deal with these cases, this solution ends up a failure because it can't account for the opposite kind of case — the case where there is a mental item but we have no conscious awareness of it. Smith gives the following argument for such a possibility:

Suppose each day on his way to work John buys a newspaper at a certain newspaper stand. However, he has to go out of his way to do so, and for no apparently good reason. The newspaper he buys is on sale at other newspaper stands on his direct route to work, there is no difference in the price or condition of the newspapers bought at the two stands, and so on. There is, however, the following difference. Behind the counter of the stand where John buys his newspaper, there are mirrors so placed that anyone who buys a newspaper there cannot help but look at himself. Let's suppose, however,

¹⁷Of course, Wright and McDowell's views on this are significantly different. Wright thinks we can specify certain conditions (the C-conditions) in which our judgements are guaranteed to be fact constituting, but McDowell need not endorse such an account. Indeed, given his overall fallibilist epistemological picture (according to which we need not be able to specify conditions in which we will be guaranteed to have accurate experiences of the outer world) it would be odd for him to specify a set of conditions, like Wright's C-conditions, in which our inner experiences would be guaranteed to be accurate. But if we are comfortable with McDowell's epistemic fallibilism there is no reason to think that this is a problem with his view.

that if it were suggested to John that the reason he buys his newspaper at that stand is that he wants to look at his own reflection, he would vehemently deny it. And it wouldn't seem to John as if he were concealing anything in doing so. However, finally, let's suppose that if the mirrors were removed from the stand, his preference for that stand would disappear.

If all this were the case, wouldn't it be plausible to suppose that John in fact desires to buy his newspaper at a stand where he can look at his own reflection; that, perhaps, he has a narcissistic tendency and buying his newspaper at that stand enables him to indulge it on the way to work? And wouldn't it also be plausible to suppose that he does not believe that this is so, given his, from his point of view, sincere denials? If this is agreed, then we have reason to reject the principle left to right: that is, we have reason to deny that if a subject desires to ϕ then she believes she desires to ϕ . (Smith, 1994, p. 106).

Smith concludes that any acceptable conception of the epistemology of desire must "allow that subjects may be *fallible* about the desires they have" (Smith, 1994, p. 107), i.e. there must be a distinction between it seeming to a subject that they desire to ϕ and it actually being the case that they desire to ϕ . That alone isn't problematic for McDowell, since, as we've just seen, it seems he may be able to make sense of the idea of misidentifying a mental item when having an experience of it. But we can draw a stronger moral from Smith's argument here, it's not merely that we make mistakes when identifying our desires, it's also possible for us to have a desire with no associated inner experience of it at all. John has no awareness at all of his desire to see himself in the mirror, he doesn't have a false experience of a desire and then identify it incorrectly.

But if, as McDowell claims, inner objects "have no existence independently of [the awareness of "inner experience"]" (McDowell, 1994/1996, p. 21), and they "have no being independently of that fact that the concepts they instantiate figure in the content of consciousness" (McDowell, 1991/1998f, p. 311), then how is this possible? If John has not had the experience of desiring to look at himself in the mirror then it seems that McDowell is forced to claim that he hasn't formed that desire at all.

And it seems that this problem will recur in the case of meaning. We can respond to questions unthinkingly, with no inner experience of any particular meanings coming to mind; but it would be strange to say that such sentences have no meaning because we were not thinking about what we were saying. Inner experience doesn't seem necessary to mean things by our words. Such cases are less clear than cases involving intentions which we were unaware of, but they at least seem possible.

So, it seems we must reject McDowell's claim that inner objects "have no existence independently of [the awareness of "inner experience"]" (McDowell, 1994/1996, p. 21), in order to make sense of the possibility of people having desires of which they have no conscious awareness. This is problematic because this claim is essential for how McDowell draws the distinction between the inner and outer worlds, so it seems that he needs

to develop a way to distinguish between inner and outer which is compatible with the possibility of desires of which we know nothing. Also, if the facts about our intentions and meanings are not constituted by our experience of them then it's not clear what does constitute them.

However, if we ignore the problem with drawing the distinction between the inner and the outer, and simply reject McDowell's claim that inner objects can only exist if we have awareness of them then it is clear that his account can make sense of the seems right / is right distinction. The facts which determine meaning are independent of what we, or anyone in our community, happen to think they are. Even in cases where no one has currently realised that we're using a concept incorrectly we can still make sense of the possibility of error, since all concepts are necessarily thinkable from a second or third-person perspective. Similarly, we can draw the seems right / is right distinction with respect to the question of which meaning we have in mind because we can make sense of the possibility of someone else (or ourselves at a future time) recognising that we've made a mistake about what meaning we have in mind.

3.10 Conclusion

Like Wright, McDowell endorses a non-reductive account of meaning. Wright does this by claiming that meaning is judgement-dependent, and that the facts about meaning are *constituted* by what we'd judge in best conditions. McDowell claims that no such account is required, and instead argues that if we simply reject the master thesis — the claim that mental items require interpretation to be normative / have meaning — then any worries about how meaning is possible will disappear. He argues that not only is the master thesis strongly counter intuitive, but he uses arguments along the lines of Kripke's sceptical argument to show that if we accept the master thesis then we are forced into a devastating infinite regress.

If we reject the master thesis this means that we accept that things can come to mind that are capable of “[imposing] a sorting of extra-mental items into those that accord with it and those that do not” (McDowell, 1993/1998h, p. 270). In other words, normative items like concepts can come to mind, and our inner experience is conceptually structured. If we add to this McDowell's claim that our experience of the outer world is also conceptually structured then it becomes clear how we can understand other people's meanings: we can simply “hear someone else's meaning in his words” (McDowell, 1984/1998o, p. 253) — when we hear someone saying ‘68 + 57 is 125’ we have a conceptually structured experience that they are saying *that 68 plus 57 is 125*. Once we've endorsed this view of our inner and outer experience then there is nothing more that needs to be done to account for the possibility of meaning.

At the start of this chapter I gave the following list of conclusions I'd reached about

straight solutions to the problem of meaning:

- We need to give an account of a fact that constitutes having a meaning in mind which is normatively significant and determines the correct extension of our expressions.
- We need to be able to draw the seems right / is right distinction, both for individuals and for whole communities.
- Meaning could be non-reductive; the fact which constitutes meaning could be known non-inferentially.
- If meaning is non-reductive then we need some kind of account of it, and judgement-dependent accounts seem to fail.

Because McDowell endorses a non-reductive account of the facts which constitute meaning, when asked “What fact makes it the case that Jones means plus by ‘+’?” his answer can be simply *the fact that Jones means plus by ‘+’*. The falsity of the master thesis shows that Jones is able to know what he means by his inner experience of this fact, and other speakers are able to know this by hearing *that Jones means plus* when he says the word ‘plus’.

Like Wright, McDowell endorses a non-reductive account of meaning, and claims that we can have non-inferential knowledge of our meanings. So he agrees with the claim of the third bullet point. He would, however, reject the fourth bullet point. *Contra* Wright, McDowell doesn’t think that we need to give an account of how non-reductive meanings are possible. Instead he endorses a quietist approach and claims that all we need to do is undermine the source of the worries that non-reductive knowledge of what we mean might not be possible. He hopes to do this by pointing out that the master thesis is common sense, and by showing how inner experience fits into our general picture of our experience of the world; a picture of experience which is required if knowledge is to be possible at all.

Things get more difficult for the second bullet point. McDowell can draw the seems right / is right distinction for the application of our concepts simply by pointing out that we can make sense of making mistakes in judging what a particular concept requires of us. We can make sense of this possibility because we understand what it is to recognise (from a second or third-person perspective) someone making a mistake in their application of a concept. And he can run precisely the same line for the mistakes of whole communities.

This solution was problematic because when looking at how McDowell draws the seems right / is right distinction it became clear that his account, at least as it currently stands, cannot account for the possibility of there being inner items of which we lack any conscious experience. According to his view, inner objects have no being independently of our experience of them, and so it seems that there can be no inner items of which

we have no awareness. This is problematic because it seems hard to deny that this happens. As Smith demonstrated, it seems reasonable to think that someone's behaviour can be guided by an intention of which they have no conscious awareness. It seems that someone can have inner states of which they themselves have no awareness, and so our inner experience cannot be all that there is to our inner states. And so we should reject McDowell's claim that they "have no existence independently of [the awareness of "inner experience"]" (McDowell, 1994/1996, p. 21). But then we are left with the problem of saying what does constitute our inner states, and of drawing the distinction between the inner and the outer worlds.

Chapter 4

McDowell's moral realism part 1: Mackie's error theory

McDowell's moral realism is perhaps one of the most interesting aspects of his philosophy. By making essential use of a Neurathian account of justification, and by drawing on his overall view of how training enables us to perceive certain aspects of the world (such as how learning a language enables us to “hear someone else's meaning in his words” (McDowell, 1984/1998o, p. 253)), McDowell develops an account of how we are able to perceive moral facts which doesn't just rely on the lazy analogy with visual perception of which moral intuitionism is often accused.

McDowell's account of his view comes out of his engagement with Mackie and Blackburn's alternatives to realism, and so I shall explain his view in the context of these two debates. Mackie argues that there are no moral facts, because if there were such facts they'd be metaphysically and epistemologically queer, and so all positive atomic moral sentences must be false. Like McDowell, Blackburn attempts to earn our right truth in ethics, but he thinks that the only way we can do so is from a metaphysically austere starting point which doesn't involve moral properties. In the first of these chapters I shall look at McDowell's response to Mackie's arguments against moral realism, and then in the second I shall turn to his response to Blackburn.¹

I shall explain McDowell's response to Mackie by drawing heavily on his “no-priority” view. This is a view which doesn't become explicitly named by McDowell until his engagement with Blackburn in “Projection and Truth in Ethics” (McDowell, 1987/1998l), but I shall argue that we can use it to best make sense of his responses to Mackie in the papers

¹I should note that in these chapters I am not intending to refute all forms of error theory and non-cognitivism. Since Mackie and Blackburn's work a number of new refinements in these meta-ethical positions have come on the scene. Moral fictionalists, such as Joyce in *The Myth of Morality* (Joyce, 2001), have developed refinements of error theory, and positions like Gibbard's norm-expressionism in *Wise Choices, Apt Feelings* (Gibbard, 1990) have suggested alternative non-cognitivist accounts. My goal here is simply to explain McDowell's own view by evaluating his criticisms of Mackie and Blackburn, not to show that no view along the lines of Mackie's and Blackburn's could be correct.

“Aesthetic Value, Objectivity, and the Fabric of the World” (McDowell, 1983/1998a), and “Values and Secondary Qualities” (McDowell, 1985/1998m). In those papers he effectively argues that the no-priority view is true of secondary qualities like colour, as well as values, and his chief arguments in those papers involve his drawing an analogy between values and secondary qualities to demonstrate that values are not in fact as queer as Mackie makes them out to be. A key part of McDowell’s argument for the reality of colours and values is his argument against Williams’ absolute conception — a conception of genuine reality as containing only things which can be understood independently of how they seem to certain perceivers — and so I shall start by explaining his argument against the absolute conception, before turning to look at his response to each of Mackie’s arguments against moral realism in turn. But before getting into McDowell’s responses to Mackie, I shall start by looking at Mackie’s own view.

4.1 Mackie’s error theory

In *Ethics: Inventing Right and Wrong* Mackie (1977) argues for a cognitivist error theory about values. What this means is that he accepts the typical moral realist’s analysis of our moral language, only he rejects the key ontological claim of moral realism — he accepts that when we say that something is good or bad we are making a claim about a fact that is supposed to be inherently motivating, but he argues that there are no such facts. This leads him to conclude that all atomic moral sentences are false — all atomic moral sentences make claims about a moral fact, but there are no such facts and so all such sentences must be false. Although he focuses particularly upon the denial of moral value judgements, he notes that he intends his arguments to generalise to all values, including aesthetic values.

Mackie claims that the correct understanding of our moral concepts is a cognitivist one — when making a moral statement such as ‘stealing is wrong’ you are making a factual claim that can be true or false. In contrast, non-cognitivist views of morality hold that such statements do not actually make a truth-apt claim, instead they hold that such claims express non-cognitive states such as desires or emotions. So, according to the non-cognitivist, by saying that ‘stealing is wrong’ you are not making a claim about a fact, but are actually expressing your disapproval of stealing. This means that even if there is no fact which makes it the case that stealing is wrong there may be nothing wrong with saying that ‘stealing is wrong’, since by saying this you are merely expressing a negative attitude towards stealing. Mackie argues that non-cognitivism is wrong as an analysis of our ordinary understanding of moral concepts because it can’t account for the apparent authority of ethics. When making moral claims it seems to us that we are not merely expressing our own attitudes, but are making a claim about something external

to ourselves which we are attempting to have knowledge of.²

This means that moral claims such as ‘stealing is wrong’, if they are true, must be made true by facts which are independent of us. But as Mackie points out, this analysis of our language isn’t enough to show that moral realism is correct: “The claim to objectivity, however ingrained in our language and thought, is not self-validating. It can and should be questioned.” (Mackie, 1977, p. 35). Mackie’s cognitivist analysis of our moral concepts holds that the truth of moral claims depends upon there being external facts which make them true, but he goes on to argue that there are no such facts and so we should adopt an error theory about value and claim that all atomic moral sentences are false:

the denial of objective values will have to be put forward not as the result of an analytic approach, but as an ‘error theory’, a theory that although most people in making moral judgements implicitly claim, among other things, to be pointing to something objectively prescriptive, these claims are all false. (Mackie, 1977, p. 35).

It is important to note that the values which Mackie intends to deny are those which involve categorical imperatives, and not hypothetical imperatives. Hypothetical imperatives are ought statements whose normative force is contingent upon a particular desire. For example, ‘If you want to be healthy, then you should eat lots of greens’ is a hypothetical imperative because its normative requirement — that you *should* eat lots of greens — only applies if you have the desire to be healthy. In contrast, categorical imperatives such as ‘You should not steal’, make unconditional ought claims: their normative force is supposed to be independent of any particular desires. If it is true that ‘You should not steal’, then regardless of whatever desires you have it still has the normative force in claiming that you should not steal. Mackie has nothing against values which involve only hypothetical imperatives, and instead intends his error theory only to reject the truth of value claims which involve categorical imperatives:

So far as ethics is concerned, my thesis that there are no objective values is specifically the denial that any such categorically imperative element is objectively valid. The objective values which I am denying would be action-directing absolutely, not contingently (in the way indicated) upon the agent’s desires and inclinations. (Mackie, 1977, p. 29).

But if there are no moral facts, then why does it seem to us that things are good or bad? Mackie explains our apparent experience of moral facts by endorsing a Humean

²Of course, the debate between cognitivism and non-cognitivism in ethics is still very much a live one, but for now I’m going to simply grant Mackie’s claim here, since it’s not particularly relevant for my purposes. As I said at the beginning of this chapter, I’m using Mackie as an example of a cognitivist rejection of moral realism, in order to explain McDowell’s position in the light of how he responds to Mackie’s. In the next chapter I’ll turn to Blackburn to provide an example of a non-cognitivist account of moral judgements (although, as we’ll see in the next chapter, there are problems with calling Blackburn a non-cognitivist).

projectivism.³ He argues that when we seem to experience something as good or bad this is not a result of us observing the moral properties of items in the world, but is instead a result of us reading our own feelings into them. He claims that:

We get the notion of something's being objectively good, or having intrinsic value, by reversing the direction of dependence here, by making the desire depend upon the goodness, instead of the goodness on the desire. (Mackie, 1977, p. 43).

Mackie also claims that, in addition to sometimes being a result of our own feelings and desires, our moral attitudes also result from internalising social demands, and this enables society to function (Mackie, 1977, pp. 42–46).

It's important to note that holding that our desires and feelings play a part in causing our moral judgements doesn't make Mackie a non-cognitivist. Non-cognitivism is the view that our moral judgements merely *express* our sentiments, and not beliefs. Whereas Mackie holds that when we make a moral judgement, even though this may be ultimately *caused* by our desires, by doing so we express a belief — when we say that 'stealing is

³I shall use "projectivism" to describe the view that our moral experience is not caused by genuine facts about values, but is instead a result of us projecting our value judgements onto the world. This means that it is a view to which anyone who denies moral realism but accepts that we seem to have experience of values must subscribe — since they need to describe our experience in some way. From this point different anti-realists diverge: cognitivist error theorists like Mackie treat our moral judgements as making false factual claims about the world, whilst non-cognitivists like Blackburn treat our moral judgements as having more to do with attitudes than representations of the world (although as I've already noted there is significant controversy over describing Blackburn as a non-cognitivist; I'll leave exploration of this until the next chapter). Since McDowell describes both Mackie and Blackburn as projectivists it seems that my use of the term follows his.

It is important to highlight how I'm using the term "projectivism" because others, such as Miller (and possibly Blackburn), use it to refer exclusively to the non-cognitivist side of these views. Miller describes projectivism as the view that "what we are doing when we treat wrongness as a genuine feature of things is *project our sentiments or emotions on to the world*" (Miller, 2013, p. 37, emphasis in original), and states that "Projectivism thus differs from cognitivism, which has to posit a realm of distinctively moral facts, as well as a mechanism which accounts for our awareness of those facts." (Miller, 2013, p. 49). Using this definition Mackie would not count as a projectivist, since he is a cognitivist and thinks that moral judgements are truth-apt and do not primarily involve non-cognitive states like sentiments or emotions.

Blackburn's use of the term is more ambiguous; he sometimes seems to use it to describe Mackie's position, but at others he explicitly describes it as the view that moral experience involves the projection of our own sentiments, and not the projection of cognitive attitudes like beliefs. For example he says:

Let us call the Humean picture of the nature of morality, and the metaphysics of the issue, projectivism. On this view we have sentiments and other reactions caused by natural features of things, and we 'guild or stain' the world by describing it as if it contained features answering to these sentiments, in the way that the niceness of an ice cream answers to the pleasure it gives us. (Blackburn, 1985/1993b, p. 152).

(For more on Blackburn's use of the word "projectivism" pages 184–185 of his paper "Attitudes and Contents" (Blackburn, 1988/1993a) are particularly useful.)

To avoid any potential confusion I'm simply going to use "projectivism" to describe the view that our moral experience involves a projection of values onto the world, and not the detection of moral facts. Therefore it is neutral on whether we should understand our moral judgements to be expressing beliefs or desires.

wrong' we are expressing the belief that it is a fact that stealing is wrong. This judgement is truth-apt, and its truth depends upon whether or not that fact obtains; it doesn't matter if we were ultimately caused to make this particular judgement because of our non-cognitive attitudes.

4.1.1 Mackie's arguments for error theory

The argument from relativity

Mackie argues for endorsing an error theory about moral values by attempting to show that if there were moral properties they'd be epistemically and metaphysically "queer", and by using the relativity of moral judgements. The argument from relativity starts by pointing out the significant variation in moral codes between times, societies, and between groups and classes within a community:

The argument from relativity has as its premise the well-known variation in moral codes from one society to another and from one period to another, and also the differences in moral beliefs between different groups and classes within a complex community. (Mackie, 1977, p. 36).

He argues that "radical differences between first order moral judgements make it difficult to treat those judgements as apprehensions of objective truths" (Mackie, 1977, p. 36). This is because were moral realism true, and we knew moral truths through the perception of moral reality, then we would expect there to be general agreement in what the moral truths are. But instead we have widespread, often quite radical, moral disagreements. Mackie argues our moral values seem to be a reflection of our involvement in particular social groups and our way of life, rather than a result of moral perception:

Disagreement about moral codes seems to reflect people's adherence to and participation in different ways of life. The causal connection seems to be mainly that way round: it is that people approve of monogamy because they participate in a monogamous way of life rather than that they participate in a monogamous way of life because they approve of monogamy. (Mackie, 1977, p. 36).

At its heart the argument from relativity is an argument to the best explanation; it is the argument that we should endorse a projectivist error theory, rather than moral realism, because it best explains the fact that there is widespread moral disagreement:

In short, the argument from relativity has some force simply because the actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values. (Mackie, 1977, p. 37).

If moral realism were true, and we knew moral facts via some kind of perception, then it would be strange indeed for us to have the widely differing moral beliefs we do, and even stranger for social groups to tend to share beliefs which aren't generally shared by the rest of humanity. In contrast, projectivism about values can much more readily explain this phenomenon, since it is the view that our apparent moral perception is a result of us projecting moral properties out onto the world, and which properties we are likely to project is in part a result of our membership in certain social groups.

Of course, disagreement alone isn't enough to undermine objectivity. Mackie points out there is disagreement in the sciences and in history too. But he argues that scientific disagreement is best explained as being a result of "speculative inferences or explanatory hypotheses based on inadequate evidence, and it is hardly plausible to interpret moral disagreement in the same way" (Mackie, 1977, p. 36) — we can explain disagreement over scientific facts as being a result of the limitations in our evidence and the different theories scientists are attempting to use to explain such evidence. Mackie claims that there is no easy parallel epistemological story we can tell to explain moral disagreement, or at least any story we can tell seems less suited to explain the facts of disagreement than simple projectivism.

Another potential problem for Mackie's argument here is the fact that moral judgements are not purely matters of convention; there is disagreement between members of the same social groups and there have always been moral heretics and reformers. If this is the case then can we simply explain our moral beliefs as caused by our membership in certain social groups? Mackie responds to this worry by arguing that we can usually understand such disagreement as the extension of moral rules which were already part of their existing way of life, "new and unconventional changes [which] seemed to them to be required for consistency" (Mackie, 1977, p. 36). Moral heretics tend to agree with the bulk of their society's moral judgements, but argue for changes because of current inconsistencies among their society's current moral beliefs. And so the existence of moral heretics does not undermine Mackie's claim that membership in a social group is the root cause of an individual's moral beliefs.

The arguments from queerness

Mackie's argument from queerness has both metaphysical and epistemological forms. The metaphysical argument aims to show that if there were objective moral values then they would be very strange: "they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe" (Mackie, 1977, p. 38). The epistemological issue essentially involves Mackie claiming that all moral realists are committed to some "special faculty of moral perception or intuition" (Mackie, 1977, p. 38) which is required to know what the moral facts are, and he argues that this would be

queer because it is “utterly different from our ordinary ways of knowing anything else” (Mackie, 1977, p. 38). And finally, he argues that even if we accept that there is such a special faculty it doesn’t explain our ability to know the strange relation between the natural and moral properties. I’ll start by looking at his metaphysical argument, before turning to look at the epistemology.

Mackie’s metaphysical argument for the queerness of moral properties is focused on the fact that moral properties “would have to be intrinsically action-guiding and motivating” (Mackie, 1977, p. 49). To bring this out he emphasises the features of Plato’s view which would have to be part of any moral realism which included categorical imperatives:

Plato’s Forms give a dramatic picture of what objective values would have to be. The Form of the Good is such that knowledge of it provides the knower with both a direction and an overriding motive; something’s being good both tells the person who knows this to pursue it and makes him pursue it. An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. (Mackie, 1977, p. 40).

To justify the “queerness” of the categorically motivating nature of moral facts Mackie uses Hume’s is-ought principle. The is-ought principle is the claim that you can’t get an ought from an is, i.e. the claim that you can never infer the truth of an ought-judgement — a judgement about how one should act — from premises involving only descriptions of how the world is.⁴ This principle amounts to the claim that natural entities can’t be intrinsically motivating and, since moral facts would be intrinsically motivating, if it is true then there can’t be any natural moral facts.⁵

However, Mackie doesn’t simply presume the truth of Hume’s principle (that would make the argument against this form of moral realism trivial) but instead claims that none of the “ordinary objects of knowledge” (Mackie, 1977, p. 40) — items in the natural world such as tables and chairs — are categorically motivating, and so the value-facts required by the moral realist must be “of quite a different order from anything else with which we are acquainted” (Mackie, 1977, p. 40). For example, consider the fact that there

⁴See *A Treatise of Human Nature* (Hume, 1739/1969, Book 3, §1.1).

⁵It’s important to note that Mackie’s argument against moral realism here relies on the claim that moral facts provide categorical reasons for action and are intrinsically motivating — “The Form of the Good is such that knowledge of it provides the knower with both a direction and an overriding motive” (Mackie, 1977, p. 40). If moral imperatives were only hypothetical imperatives then the is-ought principle would not be problematic for the moral realist because the motivating power of value facts could always be explained as the result of being combined with a desire (such as the desire to be a good person). This means that Mackie’s argument would have no force against a moral realist who endorsed externalism about moral motivation (the denial that moral values are intrinsically motivating), such as Railton (1986) and Brink (1989). However, this point is not significant here since McDowell agrees with the view of moral reasons which Mackie’s argument requires.

is a table in the room; such a fact itself is not itself intrinsically motivating. It is only when the combined with a desire, such as a desire to remove all tables from the room, would one be motivated to act and remove the table. The table alone doesn't motivate us to act. Mackie's aim at this point is not to prove that it is *impossible* for there to be facts which are intrinsically motivating, but simply to show that such facts would be metaphysically queer, and if nothing else in reality is like moral facts in this way then we should avoid endorsing views which require them. This means that if a projectivist error theory can explain our moral judgements without appeal to such facts then it should be preferred to moral realism.

Mackie's argument for the epistemic queerness of moral realism goes as follows:

Intuitionism has long been out of favour, and it is easy to point out its implausibilities. What is not so often stressed, but is more important, is that the central thesis of intuitionism is one to which any objectivist view of values is in the end committed: intuitionism merely makes plain what other forms of objectivism wrap up. Of course the suggestion that moral judgements are made or moral problems solved by just sitting down and having an ethical intuition is a travesty of actual moral thinking. But, however complex is the real process, it will require (if it is to yield authoritatively prescriptive conclusions) some input of this distinctive sort, either premisses or forms of argument or both. When we ask the awkward question, how can we be aware of this authoritative prescriptivity, of the truth of these distinctively ethical premisses or of the cogency of this distinctively ethical pattern of reasoning, none of our ordinary accounts of sensory perception or introspection or the framing and confirming of explanatory hypotheses or inference or logical construction or conceptual analysis, or any combination of these, will provide a satisfactory answer; 'a special sort of intuition' is a lame answer, but it is the one to which the clearheaded objectivist is compelled to resort. (Mackie, 1977, pp. 38–39).

Mackie here argues that all moral realists will be committed to the existence of the strange faculty of moral intuition — the ability to know the moral properties of an item in the world by direct observation — which Moore advocated as part of his non-natural moral realism, since whatever the moral realist's account of the epistemology of ethics is, it will ultimately have to include the use of some kind of moral intuition.⁶ If the realist attempts to avoid citing immediate moral experiences (such as Moorean intuitions) to provide justification for their moral beliefs then they will instead have to make use of arguments to justify these beliefs. But what supplies the premisses for such arguments? Mackie claims any argument with a moral conclusion will have to involve some kind of moral claims as its premisses. These moral claims will ultimately have to come from

⁶For Moore's view see *Principia Ethica* (Moore, 1903) and *Ethics* (Moore, 1921).

somewhere, namely a moral intuition, and so if there are no moral intuitions then our moral reasoning will simply be unable to get off the ground. Mackie concludes that all moral realists are ultimately committed to there being moral intuitions, and since moral intuition is such an unsatisfying and queer explanation the epistemology of our access to moral truth we should reject moral realism.

Mackie's argument here implicitly relies upon Hume's is-ought principle, which we saw Mackie make explicit use of in his metaphysical argument for queerness. If this principle is false and, *contra* Hume, you can justify an ought claim by using only descriptive premises then it would be possible to reach justified moral conclusions using arguments starting with only non-moral premises, and so Mackie's argument would fail. This would mean that there would be no need for the moral realist to make use of the dubious moral intuitions at all — they could simply argue for their moral claims using ordinary natural premises. However, the argument over the is-ought principle is longstanding and complicated, so for now I shall simply note that Mackie's argument for epistemological queerness seems to require belief in the is-ought gap. Also, this point is not significant for the debate between Mackie and McDowell, since McDowell also denies that you can infer a moral claim using an argument involving only non-moral premises. As we shall see, McDowell instead attempts to make sense of the idea of moral experience and so this in effect constitutes a direct response to Mackie's argument.

The final way that Mackie argues that moral realism would be queer is brought out by asking how moral properties are linked with the natural world:

What is the connection between the natural fact that an action is a piece of deliberate cruelty — say, causing pain just for fun — and the moral fact that it is wrong? It cannot be an entailment, a logical or semantic necessity. Yet it is not merely that the two features occur together. The wrongness must somehow be 'consequential' or 'supervenient'; it is wrong because it is a piece of deliberate cruelty. But just what *in the world* is signified by this 'because'? And how do we know the relation that it signifies, if this is something more than such actions being socially condemned, and condemned by us too, perhaps through our having absorbed attitudes from our social environment? It is not even sufficient to postulate a faculty which 'sees' the wrongness: something must be postulated which can see at once the natural features that constitute the cruelty, and the wrongness, and the mysterious consequential link between the two. Alternatively the intuition required might be the perception that wrongness is a higher order property belonging to certain natural properties; but what is this belonging of properties to other properties, and how can we discern it? How much simpler and more comprehensible the situation would be if we could replace the moral quality with some sort of subjective response which could be causally related to the detection of the natural features on

which the supposed quality is said to be consequential. (Mackie, 1977, p. 41, emphasis in original).

In this passage Mackie is referring to the supervenience of moral properties upon natural properties. One aspect of supervenience is covariance — if two things are the same in all their natural properties then they must be the same in all their moral properties, and you can't get a change in a thing's moral properties unless its natural properties have also changed. So, for example, consider the act of stealing. Most would agree that stealing for greed alone is morally wrong — stealing a television just because you'd like a bigger set is morally wrong. But if we change some other properties of an act — say someone is stealing a television because their family is starving and they intend to sell the television to buy bread — then it's possible that the act of stealing the television could be morally acceptable, or even good. But if two acts have identical non-moral properties then they must also have the same moral properties; this is covariance.

Another aspect of supervenience which Mackie refers to here is dependency, an act has its moral properties in some sense *because* of its non-moral properties. For example, as he points out, an act of deliberate cruelty is morally wrong because it is an act of deliberate cruelty.⁷

Mackie argues that these properties of the moral are queer because it's not clear what they are, or how we could be able to know about them. Firstly he asks “just what *in the world* is signified by this ‘because’?” — what does it mean for an act to have a particular moral property *because* it has the natural properties it does? And secondly he questions how we could know about such supervening properties. A morally competent person is not only able to know that particular acts are good or bad, they are able to know *why* they are good or bad. Consider the example of the person stealing a television to feed their family. Morally competent observers don't merely know that this particular act of theft is good, they know that it is good *because* it will feed the family. A good person would in effect know the nature of the supervenience relation between the act and the natural properties, in that if you change certain natural properties — such as the colour of the television — that doesn't change the act's moral properties, but if other natural properties change — such as the hunger of the family — then the moral properties of the act would change.

Mackie asks how it is we can know this. Even if we acknowledged that we have some kind of faculty of moral intuition and are able to ‘see’ moral properties, that would not explain our ability to know of how the moral properties relate to the natural properties.

⁷Supervenience is typically seen to have three key aspects: covariance, dependency, and nonreducibility (Kim, 1990/1993, p. 140). At this point Mackie does not draw attention to the nonreducibility of the moral, but his endorsement of Hume's is-ought principle effectively plays this role. If we cannot infer the moral facts from purely non-moral premises then there will be no way to reduce the moral to the non-moral.

Even if we accept that when witnessing someone stealing a television to feed their family we are somehow able to ‘see’ that the act is good, how does that perception enable us to know that if their family were well fed the moral significance of this act would be different? This means that even the postulation of a special faculty of moral intuition wouldn’t be enough to explain how we are able to know about the relationship between the moral and natural properties, and the epistemology of moral realism is even queerer than it first seemed.

This means that the moral realist is not only committed to claiming that we are able to ‘see’ the moral properties of things, they must also explain how we are able to know their relationship with the natural properties of things.

So, Mackie has presented us with four problems for moral realism: moral beliefs seem relative to social groups; moral facts would be metaphysically queer because they are both facts and intrinsically action-guiding; our ability to know moral facts is queer because it ultimately requires some special faculty of moral intuition; and the epistemology is doubly queer because we’re not only able to know whether an act is wrong but also why. Mackie argues that all four of these features of the moral are best explained by endorsing an error theory about values — all atomic moral sentences are false — and holding that apparent experience of moral facts is a result of our projecting them onto the world.

4.2 McDowell’s no-priority view

In the papers “Aesthetic Value, Objectivity, and the Fabric of the World” (McDowell, 1983/1998a), “Values and Secondary Qualities” (McDowell, 1985/1998m), and “Projection and Truth in Ethics” (McDowell, 1987/1998l) McDowell develops an account of values that draws parallels with secondary qualities to justify the claim that they are real and to show that Mackie is wrong to claim that they “are not part of the fabric of the world” (Mackie, 1977, p. 15).⁸ McDowell’s fundamental claim in these papers is that secondary qualities and values involve pairs of concepts, and understanding either one of the relevant concepts in the pair requires understanding of the other. The two concepts involved are the concept of our subjective response to the facts in the world (such as the concept of the experience of something as red; what it’s like to experience something as seeming to be red), and the concept of the properties themselves (such as the concept of being red; the

⁸In fact, in the third of these papers, “Projection and Truth in Ethics” (McDowell, 1987/1998l), McDowell doesn’t justify realism by drawing parallels with secondary qualities, and his target in that paper is Blackburn rather than Mackie. But I shall be drawing on that paper in my explanation of McDowell’s response to Mackie because many of his arguments against Blackburn rely on precisely the same claims as his arguments against Mackie.

property of redness). Since neither of these concepts can be understood independently of the other neither has priority. McDowell argues that this is true of both secondary qualities (like colour), and of values, and he uses the analogy between values and secondary qualities to respond to Mackie's arguments from relativity and queerness.

One way of understanding the no-priority view is in terms of the following two distinct claims:

Understanding the response requires understanding of the properties:

This is the claim that we can't understand our experience of colours or values independently of the concept of colours or values themselves as potential properties of things in the world. For example, in the case of experience of colours McDowell claims that:

only someone who has ... a use for "... is green" can understand what it is for something to look green (McDowell, 1983/1998a, p. 123).

And he makes the same claim about value experience:

there is no comprehending the right sentiments independently of the concepts of the relevant extra features (McDowell, 1987/1998l, pp. 159–160).

Understanding the properties requires understanding of the response:

This is the claim that our grasp of the concepts of colour and value properties depends upon understanding what it's like to have a colour or value experience. Here's two versions of this claim:

[A subjective property] is one such that no adequate conception of what it is for a thing to possess it is available except in terms of how the thing would, in suitable circumstances, affect a subject — a sentient being (McDowell, 1983/1998a, p. 113).

colours figure in perceptual experience ... as essentially phenomenal characteristics of objects, qualities that could not adequately be conceived except in terms of how their possessors would look (McDowell, 1985/1998m, p. 135).

Let's flesh this out with the example of the colour green. McDowell's two claims here are as follows: firstly the claim that experience of greenness is fundamentally the experience of something in the world *seeming* to be green, and so you can't grasp the concept of the experience without the concept of the property; and secondly the claim that understanding the property of green is itself dependent upon an understanding of the phenomenal quality of green colour experience — what it is like to experience something as green. To understand either requires "understanding an interlocking complex of subjective and objective, of response and feature responded to" (McDowell, 1987/1998l, p. 166).

In each of the three papers McDowell presents this view in a different context, and argues for it in distinct ways. In “Aesthetic Value, Objectivity, and the Fabric of the World” (McDowell, 1983/1998a) he defends it by arguing against the “absolute conception of reality”⁹ (McDowell, 1983/1998a, p. 119) — a conception of reality independent of how reality seems to any particular observers — he argues that belief in the absolute conception blocks our natural belief in the no-priority view, and because we should reject the absolute conception we should endorse the no-priority view. In “Values and Secondary Qualities” (McDowell, 1985/1998m, pp. 133–141) he argues that Mackie is blind to this view of value experience because he is guilty of modelling all perception on that of primary quality perception, but that fails to account for our experience of secondary qualities. And in “Projection and Truth in Ethics” McDowell (1987/1998l) argues that critics of moral realism motivate their position by making the mistake of assuming that if moral facts are not completely independent of our subjectivity then they must be merely projected onto the world by our minds, and he presents the no-priority view as a third alternative that provides a superior account of value experience.¹⁰

In the next but one section I shall look at McDowell’s argument for the no-priority view in the context of the absolute conception, primarily focusing on the argument in “Aesthetic Value, Objectivity, and the Fabric of the World” (McDowell, 1983/1998a). I shall then turn to his account of the view in “Projection and Truth in Ethics” where he introduces the “siblings” metaphor.¹¹ And then I shall explain how McDowell’s uses his no-priority view to respond to Mackie’s arguments from relativity and queerness.

4.2.1 The two notions of objectivity

According to McDowell, Mackie’s arguments against the reality of moral facts depend upon a general presumption about the nature of reality, and what kinds of things can be real. This is the presumption that reality contains nothing which can only be understood by certain observers, and that all genuine facts are potentially understandable by any rational being. This presumption in effect denies that there is anything which fits the second side of the no-priority view — there are no properties the understanding of which requires understanding of the subjective response. This is because such properties can

⁹As we’ll see in the next section, the phrase “absolute conception of reality” comes from Williams’ book *Descartes: The Project of Pure Enquiry* (Williams, 1978, pp. 65–67, 241–245).

¹⁰As I noted above, McDowell’s target in “Projection and Truth in Ethics” (McDowell, 1987/1998l) is Blackburn’s non-cognitivist projectivism, and not Mackie. But it is useful to look at this paper when evaluating McDowell’s response to Mackie since it helps illuminate the nature of McDowell’s no-priority view.

¹¹I shall not be looking at McDowell’s argument for the no-priority view in §4 of “Values and Secondary Qualities” (McDowell, 1985/1998m, pp. 133–141), because the argument is not clear.

only be understood by beings who are able to have the relevant response, and so couldn't be understood by everyone. McDowell attempts to undermine Mackie's arguments by showing that there are in fact such properties, and so the presumption Mackie relies upon is false.

McDowell claims that what motivates Mackie's view is a conception of objective reality which parallels Williams' idea of the "absolute conception of reality" in *Descartes: The Project of Pure Enquiry* (Williams, 1978, pp. 65–67, 241–245). And so he argues against Mackie's conception of objectivity by criticising the idea of the absolute conception. In §4.3 I shall look specifically at the role Mackie's presumption about the nature of reality plays in his arguments against moral realism, but first I shall look at McDowell's argument that the presumption is wrong; the absolute conception is not a complete description of the world, and there are genuine features of reality of which the no-priority view is true.

The central mistake which McDowell accuses Mackie of is the blending of two distinct notions of objectivity. Mackie correctly demonstrates that values lack objectivity in one sense — in that they are not primary qualities — but then concludes that they lack objectivity in a different sense — that they are not real. McDowell notes that if Williams' conception of reality were correct — that all genuinely real things can potentially be understood by any rational being — then Mackie would be right to make this jump. But McDowell goes on to argue that Williams' conception of reality is fundamentally flawed and incomplete, and so this jump is unjustified. I shall start by distinguishing between these two notions of objectivity, before going on to look at Williams' absolute conception.

Objectivity as primary quality

Primary qualities are objective in the sense they are possessed by an object independently from how they will appear to us; they "are adequately conceivable otherwise than in terms of dispositions to produce suitable experiences" (McDowell, 1985/1998m, p. 139). This is because the concept of a primary quality has content that is not phenomenal, but connected to the way objects occupy space:

A conception of a specific primary quality is what it is, not simply by virtue of embodying a thought of the sensory experiences to which instances of the quality would give rise, as with a secondary quality, but crucially by virtue of the place occupied by the quality in our general understanding of the ways in which objects occupy space. (McDowell, 1989/1998g, p. 358).

For example, consider the concept of *square*. We know what it is like to have an experience of an object as square, but we can also describe a square in completely non-phenomenal terms, for example:

$[x \text{ is square} \leftrightarrow x]$ is a closed planar figure consisting of four sides of equal length all of whose interior angles are right angles (Miller, 2009, p. 441).

This means that we can understand the concept of being square independently of our ability to understand what it is to see something as square. Even someone who couldn't see could still grasp this concept, since if they understand what it is for objects to occupy space and basic geometry then they could be told this biconditional and then fully grasp the concept of *squareness*.

According to McDowell what's special about secondary qualities, such as colours, is that they lack objectivity in this sense. Secondary qualities are “distinctively phenomenal” (McDowell, 1985/1998m, p. 139) — they are “essentially phenomenal characteristics of objects, qualities that could not adequately be conceived except in terms of how their possessors would look” (McDowell, 1985/1998m, p. 135). This means that the second side of the no-priority view is true of our concepts of value and colour facts — understanding the properties requires understanding of the response. And so you can only understand what it is for something to be green — grasp the concept of *greenness* — if you understand what it is to have a green colour experience.

McDowell claims that a key aspect of secondary quality properties is that knowledge of them requires a “specific sensory apparatus” (McDowell, 1983/1998a, p. 115), they are “unintelligible except as modifications of a sensibility like ours” (McDowell, 1985/1998m, p. 143). For example, someone who is completely colour blind will not be able to know what it is for something to be, say, green.¹² In contrast, as we saw above, to conceive of primary qualities like shapes all you need is to have the a conception of objects occupying space and some basic geometry, no specific sensory apparatus is required. This leads McDowell to equate the ability to be aware of certain secondary quality facts with occupying a particular point of view — “it is natural to think of possession of the special perceptual apparatus involved in colour vision as constituting a special point of view” (McDowell, 1983/1998a, p. 118) — since the way that secondary quality facts are only available to people with the appropriate sensory capacities parallels the way that certain features of objects are only visible from certain points of view.

¹²We'll see McDowell's argument for this claim in §4.2.2, but to avoid potential confusion it's important to note that by claiming that someone who is completely colour blind wouldn't be able to know what it is for something to be green McDowell does not mean to be claiming that such a person couldn't know the *extension* of green. He thinks that a colour blind person could know and be able to predict “someone's inclination to utter the word “green” when visually confronted with something” (McDowell, 1983/1998a, p. 124, fn. 13) — therefore knowing the extension of green — but he argues that such a person wouldn't be able to grasp the sense of ‘green’ (the word for the property of greenness, not the experience) because they don't know what it is for something to look green.

Values are also subjective in this sense on McDowell's view. When drawing an analogy between moral values and fearfulness he claims that fearfulness is not "intelligible from a standpoint independent of the propensity to fear" (McDowell, 1985/1998m, p. 146). Like colour properties you can only grasp what it is for something to have the properties of good or bad if you are capable of having an experience in which something seems good or bad to you. Just as awareness of something's colour involves a phenomenal experience, values are "intrinsically such as to elicit some "attitude" or state of will from someone who becomes aware of it" (McDowell, 1985/1998m, p. 132), it is intrinsic to the concept of *good* that virtuous people will be motivated to act when they see good things. And only people who are able to know what it is like to experience something as good can understand what it is for something to be good.

Objectivity as "part of the fabric of the world"

This second notion of objectivity is the key one for the debate between realists and anti-realists about values, it is the idea of being real, factual, or "part of the fabric of the world" (Mackie, 1977, p. 15). If values lack objectivity in this sense then (presuming cognitivism about values is true) Mackie is right and we should be error theorists about values and hold that all atomic moral sentences are false, because there are no value facts out there in the world to make them true. This would mean that we must explain our seeming experience of value facts as a mere projection of our false value beliefs onto a world which actually contains no such facts.

McDowell's key claim is that things can lack objectivity in the first sense — any understanding of them requires understanding of what it is like to experience them / we cannot think about them independently from how they strike certain perceivers — but still be objective in this second sense. And if this is the case then we shouldn't be worried about values lacking objectivity in the first sense, because that is not what is crucial for their reality.

4.2.2 Williams' absolute conception

According to McDowell, Mackie's mistake is to go from showing that values are subjective in the sense that they are not primary qualities, to concluding that they are not objective in the second sense; that that they are not real. McDowell attempts to explain why Mackie may make this move by considering Williams' absolute conception of reality: he argues that if Williams' conception of reality were the right one then this move would be appropriate. He then attempts to undermine Mackie's move by criticising Williams' idea of the absolute conception, by arguing that it can't make sense of our colour experience at all.

Williams defines the absolute conception as follows:

a conception of reality corrected for the special situation or other peculiarity of various observers, and that line of thought leads eventually to a conception of the world as it is independently of the peculiarities of any observers. That, surely, must be identified with a conception which, if we are not idealists, we need: a conception of the world as it is independently of all observers. (Williams, 1978, p. 241).

The absolute conception is an attempt to transcend “local or parochial [points] of view” (McDowell, 1983/1998a, p. 117) and “to achieve a conception of things as they are in themselves” (McDowell, 1983/1998a, p. 118). And the goal of the absolute conception is to “monopolize the “reality” side of the distinction between reality and appearance” (McDowell, 1983/1998a, p. 122).

The idea here is to develop an understanding of the world as it really is, undistorted by the effects of any particular point of view. This means that it can only use “concepts which are not peculiarly ours, and not peculiarly relative to our experience.” (Williams, 1978, p. 244). Williams argues that because of this, secondary qualities should not be part of our conception of how the world really is — “it is primary and not secondary qualities that characterize the material world as it really is” (Williams, 1978, p. 246) —, because they “depend on psychological factors, are a function not just of consciousness, but of the peculiarities of individuals or species” (Williams, 1978, p. 241). As we saw above, our understanding of primary qualities like shapes is independent of the particular phenomenology involved in our experience of shapes; this means that they can be understood by any rational being with a conception of how objects occupy space, and so can safely be part of the absolute conception. In contrast, secondary quality properties like *green* come from “a distinctly human perspective” (Williams, 1978, p. 243), and by perceivers with the appropriate sensory apparatus. This means that they can only be known from certain points of view, and so Williams claims that they can’t be part of the world as it really is. However, because the absolute conception is meant to be a complete description of the entire world it can’t simply ignore secondary quality experience, since observers and their experience is part of the world (Williams, 1978, p. 245). This means that it needs to be able to explain the fact that we have experience of what seem to be colours, but do so without involving actual colour properties, since they can only be understood by certain observers.

Clearly, if this is the correct account of the world as it really is, then just by showing that a certain feature of the world is subjective in the way that secondary qualities are is sufficient for showing it can’t genuinely be real. This means that if the second side of the no-priority view is correct for the concept of a certain apparent feature of the world (i.e. our understanding of it depends on knowing what it’s like to have a certain experience) then this apparent feature of the world can’t be a genuine part of reality. According to McDowell this explains why Mackie goes from showing that values are not objective in

the first sense to concluding that they lack objectivity in the second.

McDowell's argument against the absolute conception

McDowell argues against the absolute conception by attempting to show that it can't account for our experience of colour. Since Williams' absolute conception is an attempt to be a complete description of all of genuine reality, this means it must be able to "embrace and explain all the particular points of view it transcends" (McDowell, 1983/1998a, p. 122). But McDowell argues that it can't do this for our experience of secondary qualities whilst rejecting the claim that there genuinely are such features of reality, because any successful account of secondary quality experience must involve the concept of that feature of the world. Williams wants to claim that greenness is not a genuine feature of the world — he thinks that a complete description of the world as it really is (independent of any particular point of view) would not involve the property of being green. However Williams wants to do so while continuing to be able to account for the fact that we have green phenomenal experience, so he needs to do so in a way which doesn't commit him to the idea of green objects in the world. McDowell argues that this cannot be done.

McDowell's argument starts by agreeing with Williams that the property of being green can't be part of the absolute conception — it can't be part of a conception of reality which is available to *any* rational agent whatsoever, since knowledge of the property of being green is only available to people who know what it is like for something to look green. Williams describes the perspective that makes colour experience possible as "a distinctly human perspective" (Williams, 1978, p. 243). This may not be true — there may well be other creatures with colour vision like ours — but at the very least it seems clear that colour experience is only available to creatures who possess relevant sensory capacities (i.e. eyes which can detect colour). And there could be entire species that are able to think about the world, but lack such sensory capacities. Because the property of greenness is only understandable from certain points of view Williams concludes that no objects can genuinely have the property of being green because he thinks that all genuine facts must be available independently of occupying a particular point of view.

McDowell's aim here is to argue that Williams' conclusion is wrong, and that we must consider it possible that green is a genuine property of objects in the world:

the content of the appearances to be explained in this case — how it is that things appear from the point of view in question — is not so much as intelligible except on the basis of occupying the point of view. . . . only someone who has . . . a use for "... is green" can understand what it is for something to look green (McDowell, 1983/1998a, p. 123).

According to McDowell the only way we can make sense of colour experience is as it being the experience of something in the world's seeming to be a particular colour. And

Williams too acknowledges that would be hard to do; he refers to the “discouraging task of explaining ‘... looks green’ in some way which does not presuppose any prior understanding of ‘... is green’ ” (Williams, 1978, p. 243).

If it is impossible to understand what it is for something to *look green* in terms other than those of things *seeming to be green*, and if the absolute conception doesn’t include the property of *being green*, then the absolute conception cannot claim to be a complete description of reality. As we saw above, Williams agrees with McDowell that the property of being green can’t be part of the absolute conception (a conception of reality as understandable from any point of view), but if we can’t understand green colour experience without the idea of things being green then we can’t make sense of green colour experience from the absolute conception. But it seems undeniable that people genuinely have experiences of green — experiences that involve the phenomenological quality of green, and objects seeming to be green — and so this means that there is an aspect of reality which does not fit into the absolute conception. This means that the absolute conception can’t be a complete description of reality, and so can’t be used as a test for what is genuinely real.

One potential worry with this argument is that it is irrelevant for the argument between McDowell and Mackie because it provides us with no evidence against an *error theoretic* account of the property of greenness. An error theorist can happily explain our green colour experience as the experience of things *seeming* to have the property of greenness, but then add the claim that this is a property which is not actually instantiated by anything in the real world. Since Mackie is an error theorist about moral properties, can’t he make the same move there and so avoid any pressure from this argument? In response to this worry it’s important to remember what McDowell’s goal is at this point, he is not trying to show that an error theory of colours or values *can’t* be true, instead he’s attempting to undermine a potential justification for Mackie’s support of error theory — the step from the fact that values lack objectivity in the sense that they are not primary qualities to the conclusion that values lack objectivity in the sense that they are not real. McDowell introduces Williams’ absolute conception as a possible way Mackie could justify this jump (since according to the absolute conception reality does not contain any secondary quality properties), and in the current argument he is simply aiming to show that the absolute conception is not a good conception of reality. So his goal in the current argument is simply to undermine the possible use of the absolute conception to justify the claim that values cannot be real because they are secondary qualities.

Using the example of green and green colour experience we can formalise McDowell’s argument as follows:

1. Only from a point of view which involves green colour experience can you understand what it is for something to have the propriety of being green. (Understanding the properties requires understanding of the response.)
2. The absolute conception's description of reality only includes properties which can be understood independently of unique aspects a point of view.
3. So, the absolute conception cannot include the property of being green. (From **2** and **3**.)
4. To fully understand green colour experience you need to be able to understand what it is for something to be green. (Understanding the response requires understanding of the properties.)
5. The absolute conception is a complete description of reality.
6. People have green colour experiences.
7. The absolute conception must include descriptions of green colour experience. (From **5** and **6**.)
8. The absolute conception must include the property of being green. (From **4** and **7**.)
9. Contradiction between **3** and **8**.

Premises **1** and **4** are the two aspects of McDowell's no-priority view, expressed in the case of the colour green. Premise **2** is simply a definition of the absolute conception. **6** seems undeniable. Premise **5** is a statement of the goal of the absolute conception, and, if we accept McDowell's no-priority view, is the only premise we can give up. Therefore, if we accept McDowell's claim that our ability to fully understand the property of greenness and to understand green colour experience are co-dependent then we must deny that there is a complete description of reality that can be given independent of any particular point of view.

One potential problem with this argument is that McDowell has failed to provide any argument for the two key premises; he has not demonstrated that **1** and **4** are true. He seems simply to assume that you can't fully grasp what it is to have a colour experience without understanding what it is for something to be green, and that you can't grasp what it is for something to be green without being able to have green colour experience. In other words, he simply assumes that both aspects of the no-priority view are correct.

Premise **4** seems highly plausible. A highly natural way of characterising colour experiences is of things seeming to be a particular colour, and so it seems reasonable to

conclude that you can't grasp the concept of the experience without also grasping the concept of the property.

However, premise **1** is much more problematic. Jackson famously argues for a claim along these lines in the paper "Epiphenomenal Qualia" (Jackson, 1982) using the Mary argument, in which he considers the neurophysiologist Mary who knows all the scientific facts but has lived her entire life without encountering any objects which weren't black or white. Jackson claims that, despite knowing all about the science of colour vision, when she leaves the black and white room and encounters coloured objects for the first time "she will learn something about the world and our visual experience of it" (Jackson, 1982, p. 130). So, Jackson concludes that there are facts beyond those which are part of a scientific description of reality. But this argument is highly controversial and there are many who argue that Mary, who has never had a colour experience, can still know all facts about the world. If this is the case then **1** must be false.¹³

Fortunately for McDowell it seems that at least Williams agrees, at least tentatively, with these two claims, so McDowell's argument still has significant force against him. But if someone did deny either of the two aspects of the no-priority view then McDowell would have nothing to show that they could not endorse something parallel to the absolute conception. Instead, it may be better to treat McDowell's argument here as a challenge to anyone who wants to leave secondary quality properties out of their conception of reality; if they want to do so then they must give an account of secondary quality experience in terms other than there seeming to be such facts.

If McDowell's view of colours is correct, and they fit the no-priority view, then McDowell has successfully shown that the absolute conception is not a complete description of reality. This means that merely showing that values can only be understood with reference to specific perceivers is not enough to show that they are not real, and, as we shall see, this has significant consequences for Mackie's arguments against moral realism.

It's important to note, as I did briefly in a footnote in §4.2.1, that McDowell is not claiming that from the absolute conception we wouldn't be able to know the *extension* of the property of green. As McDowell himself clarifies in a footnote:

I am not suggesting that there is a problem with the absolute conception's capacity to encompass, as a possible explicandum, someone's inclination to utter the word "green" when visually confronted with something. (McDowell, 1983/1998a, p. 124, fn. 13).

¹³For details on Jackson's argument, and the various argument see "Qualia: The Knowledge Argument" (Nida-Rmelin, 2010).

Clearly, if from the absolute conception it was possible to predict when someone would correctly use the word ‘green’ to assert that something is green that would mean that the extension of the property of greenness could be knowable from the absolute conception. In other words, someone who didn’t have the appropriate sensory apparatus (such as someone who was completely colour blind) would be able to know when a fully sighted person would use the word ‘green’, and hence know the extension of *green*.

Instead, McDowell’s key claim here is that merely knowing the extension of *green* isn’t sufficient for a full understanding of green colour experience or the property of being green. His argument here takes place at the level of sense rather than reference. Even though the person who was colour blind could know when a sighted person should utter the word ‘green’ that doesn’t mean they grasp what it is like to have a green experience. And if the no-priority view is correct, and our understanding the property of greenness is dependent upon our understanding of what it’s like to have a green colour experience, then colour-blind individuals can’t grasp the full sense of ‘green’ (the word for the property of greenness).

4.2.3 McDowell’s siblings analogy

McDowell’s no-priority view also comes up in the paper “Projection and Truth in Ethics” (McDowell, 1987/1998), where he uses it to claim that Blackburn’s argument for projectivism depends upon ignoring the possibility of a no-priority view of the metaphysics of value. I’ll go into McDowell’s criticism of Blackburn’s projectivism in much greater detail in the next chapter, but here’s a brief sketch of one aspect of the argument to show how it fits into the no-priority view.

According to McDowell, Blackburn makes the mistake of presuming that there are only two options for explaining the metaphysical underpinnings of ethical talk:

Values as the parents of our sentiments: Moral facts are in the world, completely independently from our subjectivity. We learn about them in an analogous way to how we learn about the world using (primary quality¹⁴) perception, such as by using some special moral sense, or moral intuition.

Values as the children of our sentiments: There are no real moral facts in the world, the appearance of moral facts is merely projected onto the world by our own minds.

McDowell agrees that the first of these views is unacceptable, he criticises the lack of any “detailed account” (McDowell, 1987/1998, p. 154) of moral intuition, and states that

¹⁴McDowell does not talk about the distinction between primary and secondary quality perception in “Projection and Truth in Ethics” (McDowell, 1987/1998), but it seems reasonable to infer from his other work that it is the analogy with *primary* quality perception which he thinks makes simplistic moral intuition unpalatable.

“The assimilation to the senses gives this intuitionistic position the superficial appearance of offering an epistemology of our access to evaluative truth, but there is no substance behind this appearance.” (McDowell, 1987/1998l, p. 154).

However, he argues that Blackburn is wrong to conclude that the failure of intuitionistic pictures of moral epistemology forces us to be projectivists. Instead, he argues:

But why do we have to limit ourselves to those two options? What about a position that says that the extra features are neither parents nor children of our sentiments, but — if we must find an apt metaphor from the field of kinship relations — siblings? ... Denying that the extra features are prior to the relevant sentiments, such a view distances itself from the idea that they belong, mysteriously, in a reality that is wholly independent of our subjectivity and set over against it. It does not follow that the sentiments have priority. If there is no comprehending the right sentiments independently of the concepts of the relevant extra features, a no-priority view is surely indicated. (McDowell, 1987/1998l, pp. 159–160).

This view of the moral features of reality as the “siblings” of our sentiments is the no-priority view. He accepts, with Blackburn, that the moral features of the world are not prior to our sentiments — understanding the properties requires understanding of the response — but claims that the sentiments also cannot be understood independently of the concepts of moral features — understanding the response requires understanding of the properties. And so we shouldn’t be projectivists, but instead endorse a no-priority view.

4.3 McDowell’s response to Mackie’s arguments against moral realism

4.3.1 The argument from relativity

As we saw in §4.1.1, Mackie’s argument from relativity starts by pointing out the significant variation in our moral beliefs, and the fact that moral beliefs tend to be shared by members of the same social group. Variation in beliefs doesn’t automatically undermine realism, for example there is disagreement over some scientific claims, but this can be explained as a result of the limitations in our evidence and of the different theories scientists attempt to use to explain such evidence. But Mackie claims that in the case of ethics there is no parallel epistemological story we can tell to explain moral disagreement. Instead, he argues that the best explanation of moral disagreement is provided by the rejection of moral realism and the endorsement of projectivism — moral beliefs are projected onto the world, partly because of social pressures on us — since only projectivism is able to account for why our membership in social groups plays a role in determining our moral beliefs.

McDowell responds to Mackie's argument from relativity in "Aesthetic Value, Objectivity, and the Fabric of the World" (McDowell, 1983/1998a, pp. 114–115) by attempting to provide a realist explanation of the influence our membership in social groups has on our value judgements, and for the variation in our value judgements. However, in this particular paper McDowell is responding to Mackie's argument against realism about *aesthetic* values, not moral ones.¹⁵ I shall begin by explaining McDowell's argument in this paper, but then suggest some worries about whether this can be used to defend moral values against the threat from relativity. In order to save moral values from the threat of relativity I shall instead use McDowell's arguments in "Projection and Truth in Ethics" (McDowell, 1987/1998b), despite the fact that these were originally intended as a reply to Blackburn.

McDowell's response to Mackie in "Aesthetic Value, Objectivity, and the Fabric of the World" centres around attempting to show that the influence our membership in social groups has on our value judgements is not problematic. And the only reason this influence seems problematic is because of the mistake Mackie makes about objectivity which we saw above — Mackie jumps from the fact that values are subjective in the way secondary qualities are to the conclusion that they are not "part of the fabric of the world" (Mackie, 1977, p. 15). Mackie's mistake is to think that "the world is fully describable in terms of properties that can be understood without essential reference to their effects on sentient beings" (McDowell, 1983/1998a, p. 114).

According to McDowell, the fact that values are subjective in the way that secondary qualities are means that they can only be understood in terms of how the world strikes certain sentient beings. To flesh out his position McDowell draws an analogy with colour perception. Only creatures which have the relevant sensory apparatus are able to detect colour facts; this means that there could be creatures which are unable to detect the facts about what colours objects are. McDowell argues that although this means that colours lack objectivity in the sense that they are not primary qualities — they can only be understood by sentient beings with the relevant perceptual capacities — this doesn't show that they are not real. This was the conclusion of McDowell's argument against the absolute conception — there can be genuine properties of the natural world which cannot be understood independently of their impact on certain perceivers. So, if colours are still real despite the fact that some rational beings are unable to think about them, can't we say the same about aesthetic values?

McDowell claims that in the case of aesthetic values our *way of life* plays a role analogous to that which colour sense plays in our ability to detect colour facts. Our way

¹⁵In *Ethics: Inventing right and wrong* Mackie claims that his arguments undermine realism about categorical values generally, not only moral ones. In "Aesthetic Value, Objectivity, and the Fabric of the World" McDowell attempts to defend aesthetic values from Mackie's arguments, in order to avoid the complications particular to moral values.

of life enables us to be sensitive to certain facts about aesthetic value in the world. And creatures with a different way of life will be sensitive to different aesthetic value facts, in the same way that creatures with different sensory capacities will be sensitive to different secondary quality facts. As our way of life changes our sensitivity changes, and so we will be able to detect different aesthetic value facts.

In effect, what McDowell is claiming here is that there are many different types of aesthetic value facts. Different communities are capable of being sensitive to different value facts (because the members of each communities has a certain way of life), and so they will make different value judgements. This doesn't undermine realism because the same is the case for colours — you need to have the appropriate sensory capacities to detect colours; you need the appropriate way of life to be aware of a particular subset of the facts about aesthetic value.

McDowell does not flesh out what he means by “way of life” (McDowell, 1983/1998a, p. 115) here, but one possible example comes from how exposure to a particular genre of music enables you to grasp more of its aesthetic value, and be able to understand what makes music in that genre good or bad. For example, many people complain of not being able to hear anything of aesthetic value in jazz music. But people who have had substantial exposure to jazz, particularly jazz musicians, report being able to hear things going on in the music that they were simply deaf to when they first encountered the genre. The person hearing jazz for the first time is like the colour blind person, they are simply blind to the aesthetic virtues of the music. But unlike the case of colour blindness, exposure to music can enable you to develop the capacity to understand new aesthetic values. This gives us a way we could explain why people make different judgement about the aesthetic value of jazz, since some people's way of life (such as being a jazz musician) better equips them to be sensitive to the appropriate values.

However, it's not clear that this response will work in the case of differences of moral opinion. In the moral case (and arguably in the aesthetic case too) we want to be able to make sense of the possibility of there being genuine disagreement between moral views, cases where different people make different claims about the moral facts but only one can be right. For example, we want to be able to make sense of claims like it being *always* wrong to torture, and this is true for everyone — it isn't something that is any less the case for people with a different way of life; anyone who doesn't believe that torture is wrong is mistaken.¹⁶ The danger with applying McDowell's response from aesthetics to

¹⁶One seeming problem with attributing this view to McDowell is that in his critical notice of Williams' *Ethics and the Limits of Philosophy* McDowell seems to endorse moral relativism. For example, he claims that “moral judgements ... can be different without necessarily competing” (McDowell, 1986). And he

the moral case is that we seem left with the kind of moral relativism in which the moral “facts” could just be a reflection of our way of life, and not a genuine constraint upon what we can correctly believe.

Fortunately, McDowell’s paper “Projection and Truth in Ethics” (McDowell, 1987/1998l) is useful here. In this paper McDowell is responding to Blackburn, and doesn’t mention Mackie’s argument from relativity, but as we shall see it provides the tools needed avoid the threat of relativism which his response to Mackie in “Aesthetic Value, Objectivity, and the Fabric of the World” seems to be vulnerable to.

McDowell responds to the threat from relativity using Blackburn’s notion of *ethical sensibilities*, which McDowell describes as “propensities to form various attitudes in response to various features of situations” (McDowell, 1987/1998l, p. 154). Given certain representations of the world as an input an ethical sensibility will issue a moral output, such as judging a particular act to be good or bad. The moral judgements you will make, and the moral experiences you will have, are partially a result of your ethical sensibility.

So to put the threat from relativity in these terms the apparent problem is the fact that members of different social groups have different ethical sensibilities — members of different social groups will make different moral judgements and have different moral experiences in the same situation. And Mackie argues the best explanation of this difference is that our ethical sensibilities don’t involve the genuine detection of moral facts, but instead merely the projection of moral facts onto the world.

McDowell argues that the variation in ethical sensibilities is only a threat to realism if there is no way to choose between the different sensibilities:

the threat to a substantial notion of truth lies in the idea that there is nothing really to choose between different sensibilities, and that any convergence is best thought of as a mere coincidence of subjectivities rather than agreement on a range of truths (McDowell, 1987/1998l, p. 160).

The mere fact that there are differing ethical sensibilities isn’t enough to justify the rejection of moral realism, instead “The threat to truth is from the thought that there is not enough substance to our conception of reasons for ethical stances.” (McDowell,

rejects the view that moral obligations are unconditional, and claims that it is wrong to think that if “it is true that it is right, or all right, for me to act in a certain way in certain circumstances, specified in non-ethical terms, then that must be true for anyone else too” (McDowell, 1986, p. 384). This is because “what is right, or all right, for someone to do is not independent of the moral shape that the world presents to him” (McDowell, 1986, p. 384). However, later in that same paragraph he states that not all differences in moral judgement will be like this, and that differences in moral opinion will sometimes be a case of competition, where only one of the views can actually be right. And over the next couple of paragraphs he proceeds to attempt to make sense of the possibility of genuine competition in much the same way as we saw in “Projection and Truth in Ethics”. So it seems that his view is that some differences in moral opinion can be unproblematic, but that doesn’t mean that there aren’t also some questions about which there is a unique right answer.

1987/1998l, p. 162). This leads him to make the following claim about what is required to earn our right to talk of truth in ethics:

Earning the notion of truth ... is a matter of arguing that we do after all have at our disposal a conception of reasons for ethical thinking that is sufficiently rich and substantial to mark off rationally induced improvements in ethical stances from alterations induced by merely manipulative persuasion. (McDowell, 1987/1998l, p. 156).

He argues that if we can recognise that other ethical sensibilities conflict with our own, and give *reasons* for endorsing one ethical sensibility over the others, then the mere fact that there is disagreement over the moral facts doesn't show that there are no moral truths. Similarly, if we recognise rational inconsistencies in our ethical sensibilities then we can work to improve them so that we make different moral judgements and have different moral experiences in the future.

So, McDowell's view here is that members of different social groups make different moral claims because they have different ethical sensibilities. And we can connect this up to his view in "Aesthetic Value, Objectivity, and the Fabric of the World" by saying that your way of life has an impact on your ethical sensibility. But your ethical sensibility isn't fixed and static, instead your sensibility can change — it can be improved by comparing it with others and with internal reflection. The mere fact other people have a different ethical sensibility isn't enough alone to show that you are both wrong, and that neither of you are aware of the moral facts. If these different sensibilities can be rationally compared, and we can give reasons for preferring one over another (i.e. if our choice in ethical sensibility isn't restricted to "merely manipulative persuasion"), then we can be justified in thinking that there are moral truths and that our own ethical sensibility may be enabling us to be aware of the moral truths.

Mackie pointed out that disagreement alone doesn't show we should reject realism, since even over scientific claims there is disagreement. But he argued that the difference between ethics and science is that only scientific disagreement can be easily explained in a way which is compatible with realism. McDowell's response here essentially gives us a way to explain disagreement in ethics, since by giving reasons to prefer one sensibility over another we are able to say what it is about a sensibility that leads users of that sensibility to go wrong. For example, consider coming across a person who had racial biases and often made bad moral judgements involving people of a certain race. We could explain this moral failing by referring to the racist aspects of their ethical sensibility, and then by giving reasons to justify our belief that racism is wrong we could justify preferring our own (presumably less racist) ethical sensibility over theirs; thus justifying our belief that our ethical sensibility is better at detecting the moral facts than theirs.

In "Projection and Truth in Ethics" McDowell does not provide any argument to demonstrate that we can always rationally justify choosing one sensibility over another,

and he notes that if we cannot do so then we should reject moral realism (McDowell, 1987/1998l, p. 162). So we shouldn't see his response to the problem of relativity as conclusively responding to the threat. Instead he has merely shown us what debates about the significance of variation in moral beliefs should focus on: whether or not we can rationally justify preferring one sensibility over others.

In summary, McDowell's argument against relativity in "Aesthetic Value, Objectivity, and the Fabric of the World" involves two claims: there can be genuine facts that can only be understood by those capable of detecting them (the conclusion of the argument against the absolute conception); and that our way of life enables us to detect a particular subset of the facts about aesthetic value. The latter claim explains why there is disagreement, because people with different ways of life will be sensitive to different aesthetic value facts. The first claim shows that this doesn't undermine realism, since there is nothing wrong with saying that there is a genuine fact which can't be detectable by all.

However, I raised the concern that applying this view to ethics would lead to relativism, in that it would never be correct to say things like "torture is wrong for everyone", since it seems there could be ways of life that disagreed. In response to this worry I used McDowell's arguments from his engagement with Blackburn in "Projection and Truth in Ethics" to show that as long as we can make sense of rationally comparing ethical sensibilities and giving reasons to prefer our own then we can justify thinking that only certain sensibilities enable access to the genuine moral truth.

4.3.2 The argument from queerness

Mackie argues that moral facts, were there such things, would be both metaphysically and epistemologically queer. He claims that if there were objective moral facts then they'd be metaphysically queer because they are intrinsically motivating — simply knowing the moral truths is supposed to motivate us to act. And they are epistemologically queer because any knowledge of the moral facts seems to ultimately require some special faculty of moral intuition, which is left completely unexplained.

In response to the challenge of metaphysical queerness, McDowell agrees that if Mackie's view of objective reality were correct then there would indeed be something queer about moral properties. As we saw above, Mackie seems to think that objective reality only contains items which, like primary qualities, "are adequately conceivable otherwise than in terms of dispositions to produce suitable experiences" (McDowell, 1985/1998m, p. 139). In other words, Mackie seems committed to thinking that the absolute conception could be a complete description of reality and there can be no things about which the no-priority view is correct. If this were true then it would indeed be strange if there

were items in reality, like values, which are not “fully intelligible otherwise than in terms of [their] characteristic human responses” (McDowell, 1983/1998a, p. 115) — and in the case of moral facts the relevant human response is that of being motivated to act in a certain way.

But, as we’ve seen above, McDowell rejects this view of objective reality. He argues that there can be items in the world which are objective in the significance sense — that of being “part of the fabric of the world” (Mackie, 1977, p. 15) — even though they are not objective in the sense that primary qualities are, since our understanding of them depends upon our understanding of our subjective response to them (i.e. *understanding the properties requires understanding of the response*). In the case of colour this means that there can be objectively real colour properties, even though these can only be fully understood by beings capable of having colour experience. This means that there can be genuine features of objective reality which cannot “be understood without essential reference to their effects on sentient beings” (McDowell, 1983/1998a, p. 114).

This enables McDowell to respond (McDowell, 1983/1998a, pp. 116–117) to the apparently queer fact that knowledge of values is intrinsically motivating. Just as we cannot understand the property of greenness without reference to the way green things typically cause appropriate perceivers to have a particular phenomenal experience, neither can we understand the property of goodness without reference to the fact that knowledge of them intrinsically causes people to be motivated to act in a certain way. McDowell’s argument against the absolute conception effectively was an argument that there can be genuine properties which are “part of the fabric of the world” that can only be understood with reference to their effects on appropriate perceivers, be this having a particular colour experience, or being motivated to act. If there is nothing queer about the fact that colour properties are intrinsically such as to cause appropriate perceivers to have a particular experience, then neither is there anything queer about the fact that moral properties are intrinsically motivating.

Mackie’s argument that moral facts are *epistemologically* queer is based on the claim that knowledge of the moral facts ultimately seems to depend upon some special faculty of moral intuition that is left completely unexplained. And to make matters worse Mackie points out that moral properties seem to depend upon natural properties in a queer way — a particular act may be wrong *because* it is an act of deliberate cruelty. This is problematic not only due to the fact that it’s just not clear what this ‘because’ means, but also since it makes the epistemological requirements even more demanding as it means that a morally competent person isn’t only somehow able to know which acts are bad, they also know *why* they are bad. So, we are left with two queer aspects of the epistemology of morality:

1. How are we able to know any moral facts at all; is it just some kind of unexplained moral intuition?
2. How are we able to know *why* particular acts are good — how are we able to know the relationship between the moral and the natural properties.

McDowell's response to these worries centres around his alternative to the intuitionistic account of moral perception. He agrees with Mackie that the intuitionistic position gives us merely "the superficial appearance of offering an epistemology of our access to evaluative truth, but there is no substance behind this appearance" (McDowell, 1987/1998l, p. 154). This is because this picture posits "special cognitive faculties [which] are vaguely assimilated to the senses, but no detailed account can be given of how they operate, such as might make it clear to us — as clear as it is in the case of the senses — how their exercise affords us access to the relevant range of circumstances" (McDowell, 1987/1998l, p. 154).

It is at this point that McDowell argues that the analogy between values and secondary qualities breaks down:

The disanalogy [between values and secondary qualities], now, is that a virtue (say) is conceived to be not merely such as to elicit the appropriate "attitude" (as a colour is merely such as to cause the appropriate experiences), but rather such as to *merit* it. (McDowell, 1985/1998m, p. 143, emphasis in original).

As we saw in the account of McDowell's response to the problem of relativity, he claims that we can understand our moral judgements and experiences in terms of being the result of an ethical sensibility. These sensibilities can differ from person to person, but because we can rationally justify endorsing one sensibility over others these differences alone doesn't undermine the idea that the activity of our ethical sensibility involves the detection of moral truths. It is in this sense that our ethical sensibilities don't merely need to "elicit" the experience of value, the items in the world need to merit the judgement that they have the moral properties they do.

In order to explain McDowell's view of the epistemology of value it will be useful to flesh out his position by turning to his description of it in *Mind and World*:

The ethical is a domain of rational requirements, which are there in any case, whether or not we are responsive to them. We are alerted to these demands by acquiring appropriate conceptual capacities. When a decent upbringing initiates us into the relevant way of thinking, our eyes are opened to the very existence of this tract of the space of reasons. Thereafter our appreciation of its detailed layout is indefinitely subject to refinement, in reflective scrutiny of our ethical thinking. We can so much as understand, let alone seek to justify, the thought that reason makes these demands on us only at a standpoint within a system of concepts and conceptions that enables us to think about such demands, that is only at a standpoint from which demands of this kind seem to be in view. (McDowell, 1994/1996, p. 82).

According to McDowell, as part of a decent upbringing we gain the ability to think about ethics; this will involve our being instilled with a particular ethical sensibility — a propensity to particular moral experiences and judgements in certain circumstances — and an understanding of how to reason about ethics. From this point on we will be able to rationally evaluate our ethical sensibility in order to improve it and improve our ability to detect the moral facts (and, as we saw in the response to metaphysical queerness, these facts are intrinsically motivating, but there is nothing queer about that). The final sentence of this passage is the claim that this system of thought can only be understood from within; only someone who's had moral training can understand what moral demands are, and so only from that point of view can our moral reasoning be justified.

McDowell describes our moral reasoning as fitting Neurath's image:

the appropriate image is Neurath's, in which a sailor overhauls his ship while it is afloat. This does not mean such reflection cannot be radical. One can find oneself called on to jettison parts of one's inherited ways of thinking; and, though this is harder to place in Neurath's image, weaknesses that reflection discloses in inherited ways of thinking can dictate the formation of new concepts and conceptions. But the essential thing is that one can reflect only from the midst of the way of thinking one is thinking about. (McDowell, 1994/1996, p. 81).

Essential to Neurath's image is the fact that "the boat cannot be put ashore for overhaul" (McDowell, 1995/2009b, p. 35); we cannot build the entire boat from scratch as we would on dry land, instead we can only make gradual changes to the boat to prevent the boat from sinking. In the case of moral reasoning what this means is that we can't justify our moral claims from non-moral foundations, but once we have a collection of moral views we can assess and improve our beliefs, only we can't question them all at once, otherwise we'd have nothing to assess them with — you can't question if your understanding of goodness is correct without making use of some of your beliefs about goodness. Or, as McDowell puts it "One reflects on one's inherited scheme of values . . . from inside the ethical way of thinking that one finds oneself with, not by contemplating it from the external standpoint of a theory about motivations built into human beings as such." (McDowell, 1995/2009b, p. 35).

The first of the challenges of epistemological queerness I identified above asked for an account of how we are able to know moral facts. If this is a demand to explain our ability to detect the moral properties, but in non-moral terms, then according to McDowell this cannot be done. Instead, what we can do is show that we are able to understand what it is to have good reasons for a moral belief — by rationally evaluating it and comparing it to the rest of our ethical sensibility. We can justify belief in our ethical sensibility by rationally comparing it to others, and if we find our sensibility to be defective then we can change and improve it. And in fact McDowell claims that if we find our sensibility

to be incurably inconsistent or defective then we would reject the whole thing, and so would cease to reason about values. So, although we can't justify this entire process from outside — we can't justify it without presuming some of what we've learnt from our moral upbringing — that doesn't mean that we can't justify it from within.

And by justifying our ethical sensibility from within we can earn our right to talk in terms of being able to perceive the moral properties of acts and items in reality. This links up with McDowell's view of our ability to “hear someone else's meaning in his words” (McDowell, 1984/1998o, p. 253), which we saw in the previous chapter. In that case it was by learning a language that we become able to perceive facts about what sentences mean. In the case of moral training McDowell claims that “When a decent upbringing initiates us into the relevant way of thinking, our eyes are opened to the very existence of this tract of the space of reasons [ethical requirements].” (McDowell, 1994/1996, p. 82). We could talk about linguistic training in the same way: by learning a language “our ears are opened to the very existence of the meaning of sentences”. In both cases our ability to perceive such facts is learnt as part of our upbringing and the perceptual capacity is fallible, but that doesn't mean that it won't also enable access to the relevant truths. The high degree of contentiousness of our moral claims would seem to indicate that our ability to perceive moral facts is more fallible than our ability to perceive meaning, but as long as we can rationally justify endorsing the particular ethical sensibility we do that should not prevent us from claiming that it does sometimes enable perception of moral reality.

In §5 of “Aesthetic Value, Objectivity, and the Fabric of the World” McDowell argues that the limitation of only being able to justify our moral outlook from within is not unique to ethical thought, but may in fact be common to all thought, including the scientific. This is because our notion of what justifies a belief in science depends upon our “particular location in the history of science”. McDowell argues that our understanding of justification and the scientific method will make essential use of notions like simplicity, but these notions can only be used “in the context of some specific beliefs” (McDowell, 1983/1998a, p. 126) (how complex something is depends upon the beliefs it must be fitted into). There is no fixed eternal scientific method, but it, like our scientific beliefs, is itself subject to constant rational scrutiny and gradual development. Or, as McDowell puts it in the later paper “Eudaimonism and Realism in Aristotle's Ethics”:

how the concepts are taken to hang together rationally — what considerations are taken to be reasons for what conclusions of inquiry — is the product of the historical evolution of a particular human institution. (McDowell, 1995/2009b, pp. 38–39).

Instead, McDowell endorses Putnam's position,¹⁷ that we should simply accept that all assessments are made from within a conceptual system, and from a particular point

¹⁷Here McDowell cites page 32 of Putnam's book *Meaning and the Moral Sciences* (Putnam, 1978).

of view — “It is pointless to chafe at the fact that what we believe is what *we* believe.” (McDowell, 1983/1998a, p. 128). All we can justify our beliefs with is further beliefs, and our experience of the world which is enabled by that conceptual system, but we shouldn’t allow that to undermine our confidence in the claim that they enable the “possession of a subject-matter largely independent of themselves” (McDowell, 1983/1998a, p. 128). We can’t have an “impersonal and ahistorical mode of access to reality” (McDowell, 1983/1998a, p. 128), but we can still have access to reality, it’s just necessarily from our own perspective.

In other words, McDowell’s claim is that for both science and ethics “we have only our own lights to go on, in trying to ensure that the considerations that we are responsive to are really reasons for thinking one thing rather than another” (McDowell, 1995/2009b, p. 38).

If this is correct, then we should not be concerned that we cannot give a full explanation of our ability to know the moral facts — we can only understand the justification for a moral claim from within the domain of our ethical thought. Towards the end of the next chapter I shall look into this aspect of McDowell’s view in more detail, and attempt to defend it by drawing a parallel with mathematics.

McDowell doesn’t explicitly respond to the last of Mackie’s arguments for queerness that I identified. This is the argument that acts and items in the world seem to have the moral properties they do *because* of their natural properties. According to Mackie this generates additional epistemological queerness because it means that even if we can somehow “see” the moral properties of things, it is not clear how the moral realist would explain our ability to know *why* an item in the world has the moral properties it does — how we know how the moral properties follow from an item’s natural properties.

Fortunately, McDowell’s understanding of moral epistemology in terms of our susceptibility to reasons provides us with precisely what we need to respond to this argument. As we’ve seen above, it is an essential aspect of McDowell’s view that we are able to rationally evaluate our ethical sensibility; we may not be able to evaluate it all at once, but we can still evaluate each part in turn, using the rest of our moral outlook to do so.

So, consider a particular moral judgement, such as the judgement that an act of theft to feed a hungry family is good. McDowell does not go into detail about what is required to evaluate the value of particular judgements, but presumably in order to evaluate it we need to evaluate the *reasons* for our judgement that this act is good. In this case that would involve recognising the fact that it is good because, for example, the hunger of the children is more important than respecting property law. And then we could go on to evaluate whether that really is the reason we judge it to be good, and if that reason really

has the moral significance it seems to. Being able to do all of this kind of critical evaluation seems to be an essential part of being able to evaluate our moral views, and so it is an essential aspect of McDowell's view that we are able to understand *why* particular acts have the moral properties they do. If we only knew which acts were good and bad, and not why then surely it would be impossible to rationally evaluate our moral judgements. If this is correct then Mackie's concern here is simply not a problem for McDowell, since a response to it is a key part of McDowell's position.

4.4 Summary

Mackie argues that we should endorse an error theory about values. He accepts that the moral realist is right to take a cognitivist understanding of our moral talk and hold that moral judgements express beliefs about moral facts, but the problem is that there are no such facts. And so he concludes that all atomic moral sentences are false. Instead we should explain our seeming experience using projectivism; our experience involves our merely projecting our moral beliefs onto the world, not the detection of moral properties. He justifies this conclusion by arguing that that projectivism better deals with the fact that moral beliefs vary between social groups, and by arguing that if there were genuine moral properties they would be both metaphysically and epistemologically queer.

McDowell's response to Mackie's argument from relativity involved drawing an analogy with colour properties. If McDowell's no-priority account of colour properties is true (which he does not present particularly in-depth arguments for, but at the very least seems to be a reasonable position) then there are genuine properties of the world which can only be understood with respect to their effect on appropriate perceivers. McDowell argues that if values are like this then we can respond to the argument from relativity and explain moral disagreement by pointing out that the impact our way of life has on our value judgements may be just like the impact the possession of certain sensory capacities has on our secondary quality judgements. Not everyone will make the same secondary quality judgements because some will lack certain sensory capacities; similarly, not everyone will make the same value judgements because people have different ways of life, but that doesn't show that they aren't detecting genuine properties. However, I pointed out that this argument seems inappropriate for moral values since it seems to involve the endorsement of moral relativism, and can't make sense of universal moral judgements. I responded to this worry by making use of McDowell's discussion with Blackburn involving ethical sensibilities, where McDowell argues that mere variation in sensibilities doesn't show that none of them have the right to claim that they involve the detection of genuine truths, as long as we can rationally justify endorsing the sensibility we do.

McDowell's response to the argument from metaphysical queerness also involved the

the no-priority view and the analogy with colours. If the no-priority view is the correct account of colours then our subjective responses to colours — having a particular experience — is an essential part of our understanding of what it is for something to be a particular colour. If colours are genuine aspects of the world, and McDowell's argument against the absolute conception attempts to demonstrate this, then there are genuine aspects of the world which can only be understood with reference to how appropriate perceivers respond to them. This opens the door to responding to Mackie's argument that values would be metaphysically queer. Mackie argued that value facts would be queer since they are such that, when detected, they are intrinsically motivating to appropriate perceivers. But McDowell has already shown that there is nothing wrong with there being genuine properties of the world which can only be understood with reference to their impact on appropriate perceivers — colours are intrinsically such as to cause us to have certain visual experience — and so we shouldn't think that values are queer just because they are intrinsically motivating.

McDowell's response to the arguments from epistemic queerness involved the endorsement of a Neurathian picture of our access to truth. If this is correct then we shouldn't expect to be able to justify our ability to access the moral truths from outside — from a perspective which doesn't presume our understanding of ethics. This doesn't prevent McDowell from saying a lot about how moral truths are justified: we can reason about our ethical sensibility in order to evaluate it, not only by comparing our ethical sensibility with others but also by looking at its internal consistency. This justifies our talking in terms of perceiving the moral facts; through a normal upbringing we gain not only the ability to reason about moral claims, but also the ability to see the moral properties of things in the world, just as our upbringing enables us to hear what speakers of our language mean when they speak. In the case of our access to moral truths McDowell's epistemological picture makes essential use of our susceptibility to reasons, and I suggested that we also use this to respond to Mackie's second argument from epistemic queerness. Mackie points out that morally competent agents are not only able to know *which* acts are good or bad, they are also able to know *why* they good or bad. But if McDowell's account of the epistemology of values is correct then responding to this problem is straightforward, since it is an essential part of rationally evaluating an ethical sensibility that we can take any particular act and evaluate our reasons for judging it to be good or bad.

Chapter 5

McDowell's moral realism part 2: Blackburn's quasi-realism

It seems that Mackie's arguments against moral realism fail. Or, at least if we accept the no-priority view of values and secondary qualities, then Williams' absolute conception fails to be a complete description of reality and we cannot use the fact that values can only be understood by certain perceivers to justify rejecting their reality. McDowell argues that we can justify our claim to perceive moral facts because we can evaluate our moral thinking from within, and give rational reasons for our views in the face of moral disagreement.

In this chapter I shall consider Blackburn's quasi-realist meta-ethics. Like McDowell, Blackburn argues that the fact that we can compare ethical sensibilities and give rational reasons for preferring one ethical sensibility over others enables us to earn our right to the notion of truth in ethics. However, unlike McDowell he attempts to do this in a way which recognises the problems with moral realism that Mackie identifies; he doesn't want to simply presume from the start that there are moral facts and that we can be sensitive to them. His quasi-realist project is to go from a metaphysically austere starting point that doesn't make these assumptions, but from there earn our right to the notion of moral truth.

I shall argue that the best way to understand the conflict between McDowell and Blackburn is in the differing priority they give to two different explanatory tests. McDowell argues that Blackburn's austere starting point view is incompatible with the denial of the reality of values. He advances two arguments, the contaminated response and disentangling arguments, to attempt to demonstrate this. I shall look at these arguments in §5.4, and ultimately conclude that they fail to undermine Blackburn's position. The second explanatory test is the one which Blackburn favours. He argues that we shouldn't be realists of McDowell's sort because the supposed value facts pull no weight in McDowell explanation of our ability to know about them. I shall draw on what I said towards the end of the last chapter to explain McDowell's response to this criticism, and finally I'll

develop an analogy with my conclusions about mathematics in chapter 1 to demonstrate that McDowell's position is more reasonable than it may first appear.

5.1 Blackburn's quasi-realism

Blackburn's quasi-realism is essentially the combination of two views:

Non-cognitivist projectivism: “we have sentiments and other reactions caused by natural features of things, and we ‘gild or stain’ the world by describing it as if it contained features answering to these sentiments, in the way that the niceness of an ice cream answers to the pleasure it gives us” (Blackburn, 1985/1993b, p. 152). (Blackburn actually rejects the label of “non-cognitivism”, I shall explain why in §5.1.1.)

Quasi-realism: the attempt to show that “projectivism is consistent with, and indeed explains, the important surface phenomena of ethics” (Blackburn, 1988, p. 362), and to save the notion of truth in ethics.

Like Mackie, Blackburn thinks that we should explain our experience of moral properties as a result of our projecting moral properties onto the world, rather than as a result of our detecting moral properties. But unlike Mackie, Blackburn's a non-cognitivist — he thinks that we should understand our moral judgements as being expressions of sentiments rather than beliefs.

The biggest difference between Mackie and Blackburn is their account of what we should say about moral statements after reaching the projectivist conclusion. As we saw in the last chapter Mackie endorses an error theory about moral talk, he thinks that every atomic moral sentence is false. In contrast, Blackburn attempts to find a way to hold that atomic moral statements can be true, despite starting with a projectivist explanation of moral experience. He does this by arguing that we should understand moral claims as the expression of attitudes, and that when they express good moral attitudes they are true. Blackburn not only makes this move to claim that first-order moral statements, such as “Murder is wrong”, can be true, he also thinks that second-order moral statements like “The fact that murder is wrong is independent of me, my responses, and my sentiments” are true.

This explains why Blackburn rejects his view being labelled as “non-cognitivist”. Although his starting point views moral judgements as expressions of desires, his goal is to explain how we can earn the right to truth in ethics. He thinks that moral judgements can genuinely be true or false, and so we can earn the right to talk of moral beliefs, even though we are not justified in taking that view from the start:

I think that naturalism demands this view of ethics, but in any case it motivates it. It does so because in this package the fundamental state of mind of one who has an ethical commitment makes natural sense. This state of mind is not located as a belief (the belief in a duty, right, value). We may end up calling it a belief, but that is after the work has been done. In fact, we may end up saying that there really are values (such as the value of honesty) and facts (such as the fact that you have a duty to your children). For in this branch of philosophy, it is not what you finish by saying, but how you manage to say it that matters. (Blackburn, 1988, p. 363).

Despite Blackburn's dislike of the term "non-cognitivism" it is essential to be able to evaluate his starting point in isolation, since if that is not a coherent position then his whole project should be rejected (this is McDowell's strategy against Blackburn, as we'll see in §5.4). Blackburn's starting point is a form of non-cognitivism and so when discussing that I shall describe it in these terms. The fact that Blackburn goes on to attempt to earn the right to the notion of truth in ethics, and ultimately reject moral non-cognitivism does not undermine the value in sometimes talking about his starting point in isolation.

I shall start by looking at Blackburn's arguments for his metaphysically austere starting point, and then turn to look at how he goes on to earn truth from there, first in his response to the Frege-Geach problem, and then by looking at how he argues that moral facts are mind independent.

5.1.1 Blackburn's arguments for non-cognitivist projectivism

Blackburn's arguments for non-cognitivism

In these arguments Blackburn attempts to show that "To have a [moral commitment] is to hold an attitude, not a belief" (Blackburn, 1984b, pp. 187–188). Classifying moral commitments and judgements as expressions of attitudes, rather than beliefs, is typically called moral non-cognitivism.¹ Blackburn is critical of his view being described as non-cognitivist because he ultimately wants to justify talking of moral beliefs and to earn our right to the notion of moral truth. But before turning to look at the subtleties of this position I shall start by explaining Blackburn's argument for associating moral judgements with attitudes.

Blackburn argues that moral judgements involve attitudes rather than beliefs because this view fits better with Hume's theory of motivation. According to the Humean theory of motivation a person's motivation for an action requires the combination of both a belief and a desire. Consider the example of someone being motivated to run to catch a bus.

¹For an introduction to the terms "cognitivism" and "non-cognitivism" in metaethics see Miller's *An Introduction to Contemporary Metaethics* (Miller, 2013, p. 3).

In this case we can imagine the combination of the belief that if they don't run then they'll miss it and the desire to catch the bus combining to motivate them to run to catch the bus. Neither the belief nor the desire alone would motivate them to do anything, but their combination leads to their being motivated to act. The Humean theory of motivation holds that all motivations are results of a belief-desire pair such as this, and that belief and desire alone can never produce any motivation to act.

So, Blackburn asks, do moral judgements express beliefs or desires? He explains the two options as follows:

if moral commitments express *attitudes*, they should function to supplement *beliefs* in the explanation of action. If they express *beliefs*, they should themselves need supplementing by mention of *desires* in a fully displayed explanation of action (Blackburn, 1984b, p. 188, emphasis in original).

So, consider the moral judgement that saving Tommy, a small child, from drowning in a lake is good. If moral judgements express beliefs then, according to the Humean theory of motivation, they would need to be supplemented with desires to provide us with the motivation to act. This means that in this case the moral judgement would be the belief that saving drowning children is good, but this would only motivate us to act, and save Tommy, if we also have a supplementary desire, such as the desire to do good. Blackburn argues that this account is problematic because "It seems to be a conceptual truth that to regard something as good is to feel the pull towards promoting or choosing it" (Blackburn, 1984b, p. 188), but if moral judgements express beliefs then they need to be combined with a supplementary desire to provide us with any motivation to act. It seems possible that someone could lack this desire (say, the desire to be good), and so someone could regard something as good without being motivated to act. This violates the seeming conceptual truth (internalism about moral motivation).²

The other option is that moral judgements express desires, this would mean that they'd need supplementing with a belief to motivate us to act. So, to return to our example of drowning Tommy, the relevant moral judgement would be something like the desire to save children from drowning, and then that would need to be combined with the belief that Tommy is a drowning child in order to motivate us to act. Blackburn argues that this accurately captures the nature of moral motivation, since he thinks that it is right to think that moral judgements need only to be supplemented by our appraisal of the current situation to motivate us to act. He concludes that moral judgements express desires rather than beliefs, and it is these desires which are projected onto the world in moral experience, leading us to think that certain actions are good or bad.

If moral judgements are the expressions of our attitudes, and not beliefs, then it is

²Blackburn (1984b, pp. 188–189) notes that this seeming conceptual truth is controversial, but I shall ignore this problem since McDowell too is an internalist about moral motivation.

clear that we must be projectivists rather than straightforward realists (although, as we shall see, Blackburn's quasi-realism is an attempt to go on from this starting point and end up endorsing a kind of realism). The projectivist picture here is the view that when we make a moral judgement we express our attitude, and it is this attitude which leads to our projecting moral properties into the world, and experiencing the world as if it has moral properties. Straightforward realism is incompatible with this view of moral judgements since our moral judgements can't involve a claim about mind-independent facts if all they are is expressions of our attitudes.

Blackburn's rejection of the label "non-cognitivism"

This view — the association of moral judgement with attitudes rather than beliefs — is typically known as non-cognitivism. However, Blackburn rejects this label:

Projectivism may seem to be automatically opposed to the view that in saying that something is good (etc.) we give voice to a real belief about it, and it is often so introduced (as labels like 'non-cognitivism' suggest). But this opposition is not automatic. (Blackburn, 1988/1993a, p. 185).³

He clarifies his position as follows:

When I say that these sentences primarily express attitudes, I have never intended to deny that they can be regarded as expressing beliefs or propositions. This opposition would be going beyond anything I embrace. But I do mean that the right way of theorizing about them identifies them, in the first instance, as expressing states of mind whose function is not to represent anything about the world. They express something more to do with attitudes, practices, emotions, or feelings arising in contemplating some kinds of conduct, with goal seeking, with insistence upon normative constraints on conduct, and they express nothing to do with representing the world. In the familiar metaphor, their 'direction of fit' with the world is active — to have the world conform to them — rather than descriptive or representational. I call someone who approves of both this contrast, and this direction of theorizing, a projectivist. (Blackburn, 1988/1993a, pp. 184–185).⁴

³As we'll see in the next quote, Blackburn uses "projectivism" to refer to the combination of two claims: that moral judgements primarily express attitudes rather than beliefs, and that these are projected onto the world by us. As I noted in the previous chapter I use the word "projectivism" to refer to only the second of these two views: the view that moral experience involves the projection of moral properties onto the world, and not the detection of them. This means that my use is neutral between cognitivism and non-cognitivism — whether you should understand moral judgements as expressing beliefs or desires — and applies equally to Mackie and Blackburn.

⁴In my explanation of Blackburn's argument for associating moral judgements with attitudes above, despite what he says in this passage, I did explain Blackburn's view in terms of denying that moral judgements express beliefs. This is because I wanted to keep Blackburn's position simple so that I could explain the argument for it, before getting into the subtleties of his quasi-realism. Also, as I quoted above, Blackburn does explicitly state in *Spreading the Word* that "To have a [moral commitment] is to hold an attitude, not a belief" (Blackburn, 1984b, pp. 187–188).

So, it seems that Blackburn's view is that moral judgements primarily express attitudes, or at least we should identify them "in the first instance" as expressing attitudes rather than beliefs. When thinking about these kinds of issues in Blackburn it's essential to distinguish between his starting point, and what he ultimately wants to end up saying. I shall explain this aspect of his view in more detail in §5.3.1 ("The significance of the quasi-realist starting point"), but the essential point to recognise is that Blackburn's quasi-realist project is to go from the starting point of what is effectively non-cognitivist projectivism and to find a way to justify talking about moral beliefs, and even moral truths. Despite the fact that when we first look at the world we can find nothing which is like a moral fact that could make our moral judgements true or false (indeed, as we've just seen, Blackburn initially argues that moral judgements express desires, not beliefs, and so are not even truth-apt) his goal is to find a way of justify saying that moral judgements express beliefs that can be true. For example, he claims that:

Subtlety with the concept of belief, or with the concept of truth or of fact, may enable the expressivist to soften this opposition. Theory may enable us to understand how a commitment with its center in the expression of subjective determinations of the mind can also function as expressing belief, or be capable of sustaining the truth predicate — properly called 'true' or 'false'. (Blackburn, 1988/1993a, p. 185).

But as I said above, it is useful to be able to evaluate Blackburn's starting point in isolation, and so when talking about his starting point I shall describe it as non-cognitivism.

Metaphysical and epistemological arguments

Clearly if the argument for moral non-cognitivism is correct then we can't hold that our moral experience is a result of mind-independent value *facts*, and we should instead endorse a projectivist account of moral experience. Blackburn provides further support for the claim that projectivism is a better explanation of our moral judgements than moral realism by arguing, like Mackie, that it provides a superior metaphysics and epistemology.

When judging whether we should be realists about a particular domain of facts Blackburn claims that we should make use of "evidence from its surroundings, or in other words from other things we think, for example about explanation, truth, fact, reality and our knowledge of it" (Blackburn, 1993d, p. 371). And we should prefer an account of ethics which enables us to "bring what we say about ethics and the rest into some working relationship with a scientific world view" (Blackburn, 1993d, p. 372) over one which doesn't connect with our scientific understanding of ourselves.

He criticises McDowell's view on the grounds that "there is no theory connecting [moral] truths to devices whereby we know about them" (Blackburn, 1985/1993b, p. 163). And like Mackie he questions how it is we are able to know through perception the truth

of general moral standards, arguing that “it is stretching things to see these general standards as perceptually formed or maintained . . . Do I have an antenna for detecting timeless property-to-value connections” (Blackburn, 1988, p. 365).

Blackburn argues that the projectivist account can provide an account of our apparent moral experience, doing so without any additional questionable ontological commitments. Like Mackie, Blackburn argues that projectivism is to be preferred for reasons of meta-physical and epistemic economy:

The projective theory intends to ask no more from the world than what we know is there — the ordinary features of things on the basis of which we make decisions about them, like or dislike them, fear them and avoid them, desire them and seek them out. It asks no more than this: a natural world and patterns of reaction to it. By contrast, a theory assimilating moral understanding to perception demands more of the world. Perception is a causal process: we perceive those features of things which are responsible for our experiences. It is uneconomical to postulate both a feature of things (the values they have) *and* a mechanism (intuition) by which we are happily aware of it. (Blackburn, 1984b, p. 182).

Blackburn argues that another advantage for projectivism is how well it fits with our scientific understanding of ourselves. He points out that we can tell an evolutionary story for why we have moral values, such as:

Animals with standing dispositions to cooperate (say) do better in terms of other needs like freedom from fleas or ability to survive failed hunting expeditions by begging meals from others. (Blackburn, 1988, pp. 363–364).

There is evolutionary pressure for us to have moral values because they cause us to cooperate which ends up increasing our chances of survival.

In addition, the introduction of the evolutionary story can also be used to explain why moral values are intrinsically motivating — an aspect of values which many anti-realists hold to be difficult for the realist to explain:

Evolutionary success may attend the animal that helps those that have helped it, but it would not attend an allegedly possible animal that thinks it ought to help but does not. In the competition for survival, it is what the animal *does* that matters. This is important, for it shows that only if values are intrinsically motivating, is a natural story of their emergence possible. (Blackburn, 1988, p. 363).

If our moral values were not intrinsically motivating then possession of them would not guarantee that we would be motivated to act in accordance with them, and so would fail to guarantee that they would be of any evolutionary use. The fact that they are

intrinsically motivating means that once we endorse a particular value we cannot help but feel its force in motivating us to act (although, of course, the moral desire can be overpowered by other stronger desires).

Blackburn argues that actual rights and duties play no role in this evolutionary explanation. The reason acting cooperatively is evolutionarily valuable is utterly independent of any possible value facts:

No right, duty, or value plays any explanatory role in this history. It is not as if the creature with a standing disposition to help those who have helped it does well *because* that is a virtue. Its being a virtue is irrelevant to evolutionary biology. There is no such naturalistically respectable explanation. (Blackburn, 1988, p. 364).

This means that even if there somehow were moral truths to be discovered there seems to be no reason why evolution would put any pressure upon us to be able to detect them. Simply projecting categorical values onto our experience of the world can do all the work that is needed, and if there were any differences between the moral truths and what there is evolutionary pressure for us to feel motivated by then there would be no reason for evolution to prefer the moral truths at all — “in connection with naturalism, the question to ask of the view is why nature should have bothered” (Blackburn, 1988, p. 365).⁵

What’s next?

So, Blackburn concludes that we should endorse a non-cognitivist projectivist explanation of our value judgements. Not only does this view fit better with the Humean theory of motivation, but it also fits better with our general scientific understanding of the world and doesn’t involve an unexplained epistemology. According to Blackburn moral judgements express desires and moral experience is a result of our projecting these desires onto the world, so that it seems⁶ to us that things in the world are good or bad independently of us. But in fact no independent moral facts play any role in the correct explanation of our moral experience. So, what should we do once we’ve concluded that projectivism is true? Can we continue to moralise? Blackburn (1985/1993b, pp. 149–153) argues that there are two options for the projectivist:

⁵Blackburn also develops a third argument against moral realism: the argument from supervenience and the ban on mixed worlds, but I shall not get into that argument here. For more on that argument see *Spreading the Word* (Blackburn, 1984b, pp. 182–187), “Moral Realism” (Blackburn, 1973/1993c), and “Supervenience Revisited” (Blackburn, 1985/1993e).

⁶As we saw a little of above, and will soon see in much greater detail, Blackburn argues against concluding that things merely *seem* to have moral value. But for now it will make explaining the projectivist aspect of his account if we simply ignore that aspect of his position.

Revisionist projectivism: once we've discovered the truth of projectivism we should conduct our moral reasoning differently (perhaps even stop making moral claims altogether).

Quasi-realism: even after we purge ourselves of the metaphysical mistakes we don't have to fundamentally change our moralising and we don't need to change our moral commitments.

Mackie is an error theorist, his endorsement of (cognitivist) projectivism leads him to conclude that all atomic moral sentences are false — they make claims about moral facts in the world, but there are no such facts, and so they must all be false. And yet despite this, the majority of his book *Ethics: Inventing Right and Wrong* (Mackie, 1977) is committed to making what seem to be normal first-order moral claims!⁷ Blackburn, like many other readers of Mackie's work, points out that “there is something fishy about holding an error theory yet continuing to moralize” (Blackburn, 1985/1993b, p. 152). And the conclusion that error theory is true doesn't even lead Mackie to moralise radically differently but he instead seems to continue just as before, just as a moral realist would. But if the discovery of the truth of projectivism doesn't lead Mackie to moralise differently then, Blackburn argues, we should conclude that there is nothing wrong with the *practice* of people who believe the old realist theory. The error in their theory does not infect their practice, and we are safe to continue with the same order first-order moral theory, because “The practice could be clipped onto either metaphysic.” (Blackburn, 1985/1993b, p. 151). Blackburn argues as follows:

it seems gratuitous to infer that there are two different activities from the fact that there are two or more different theories about the nature of the activity. It would be much more natural to say that Hume and Mackie moralize, just as ordinary people do, but with a developed and different theory about what it is that they are doing. The error theory then shrinks to the claim that most ordinary moralists have a bad theory, or at least no very good theory, about what it is to moralize, and in particular that they falsely imagine a kind of objectivity for values, obligations, and so on. This may be true, but it does not follow that the error infects the practice of moralizing, nor the concepts used in ways defined by that practice. (Blackburn, 1985/1993b, p. 150).

Blackburn develops quasi-realism as an attempt to defend continuing first-order moral practice as before, despite endorsing non-cognitivist projectivism. And, unlike Mackie, not only does he think we can moralise as before, but we can also continue to hold that moral claims such as ‘Murder is wrong’ and even ‘Murder would be wrong even if I didn't think it was’ are true.

⁷The arguments for error theory we saw in the last chapter all take place in the first chapter of Mackie's book, most of the rest of the book involves first-order moral theorising.

5.2 Quasi-realism

The aim of quasi-realism is to show that “projectivism is consistent with, and indeed explains, the important surface phenomena of ethics” (Blackburn, 1988, p. 362). As we saw briefly in §5.1.1 Blackburn’s goal is to go from the projectivist starting point but from there earn our right to the notion of moral truth. In effect, he is aiming to acknowledge many of the problems with moral realism Mackie identifies, but avoid Mackie’s error theoretic conclusion, that all atomic moral sentences are false.

The essence of Blackburn’s quasi-realist strategy is the use of the following move:

what seems like a thought that embodies a particular second-order metaphysics of morals is seen instead as a kind of thought that expresses a first-order attitude or need. (Blackburn, 1985/1993b, p. 153).

Blackburn points out that we can have moral attitudes not only towards first-order ethical claims, but also towards second-order claims about the metaphysics of morals, and towards ethical sensibilities. Moral attitudes are true if they are judged to be so by the correct first-order moral theory, and so second-order claims can be true if they are the kinds of things that a good person should say.

I shall next look at two particularly problematic cases for the quasi-realist strategy: the Frege-Geach problem — the problem for non-cognitivist understandings of moral arguments — and the problem of mind dependence — the problem of second-order sentences which make claims about the metaphysics of morals. Finally I shall turn to evaluate Blackburn’s move, and suggest that he hasn’t successfully responded to the problem of the seeming conflict between quasi-realism and non-cognitivism.

5.2.1 The Frege-Geach problem

One of the most serious objections to non-cognitivist accounts of moral value is the Frege-Geach problem (Geach 1960; 1965). Consider the following argument:

1. If tormenting the cat is bad, then getting your little brother to torment the cat is bad.
2. Tormenting the cat is bad.
3. *Therefore*, getting your little brother to torment the cat is bad.⁸

Clearly this argument is valid. The problem for the non-cognitivist is that it’s difficult for them to explain its validity, because the meaning of “Tormenting the cat is bad” seems to be different in **1** than its meaning in **2**.

⁸Adapted from Geach’s example in “Assertion” (Geach, 1965, p. 463).

Non-cognitivism is the view that when we make a moral judgement, such as that in **2**, we are expressing an attitude. When I assert that “Tormenting the cat is bad” what I am doing is expressing a negative attitude towards tormenting cats. The problem with **1** is that the phrase “Tormenting the cat is bad” is not being asserted: I could assert **1** without endorsing the antecedent of the conditional, and by asserting “If tormenting the cat is bad, then getting your little brother to torment the cat is bad.” I haven’t actually expressed any attitude towards tormenting cats.

Conditionals such as **1** are typically understood to only involve statements with truth-value, and so to make that explicit we could express it as follows:

1’. If (it is true that (tormenting the cat is bad)) then (it must be true that (getting your little brother to torment the cat is bad)).

But according to non-cognitivism the second premise does not involve the claim that it is true that tormenting the cat is bad, it merely expresses a negative attitude towards tormenting the cat, something like:

2’. Boo! to tormenting the cat.

Once we’ve made explicit how the non-cognitivist would naturally read these premises it is clear that the argument becomes invalid, since **2’** is a fundamentally different from what appears in the antecedent of **1’**. From the combination of **1’** and **2’** we can infer nothing.

Blackburn’s solves this problem by introducing the idea of an ethical sensibility (which we saw McDowell make use of in the previous chapter), which he describes as input-output functions:

We can usefully compare the ethical agent to a device whose function is to take certain inputs and deliver certain outputs. The *input* to the system is a representation, for instance of an action, or a situation, or a character, as being of a certain type, as having certain properties. The *output*, we are saying, is a certain attitude, or a pressure on attitudes, or a favouring of policies, choices, and actions. Such a device is a function from input to output: an ethical sensibility. (Blackburn, 1998, p. 5).

For example, someone who had an ethical sensibility that included the attitude “Boo! to tormenting the cat” would respond to the input of a representation of acts of tormenting cats by producing the output of a negative attitude to the acts.

As we saw in the previous chapter, ethical sensibilities can be good and bad, and can be assessed. When assessing someone’s moral views as good or bad we are effectively assessing their ethical sensibility. We may reject ethical sensibilities because people who endorse them would behave horrendously, but that’s not the only way to assess sensibilities. It is also important to evaluate sensibilities based on the interactions between the

attitudes they involve. We should reject sensibilities which involve conflicting attitudes, and, crucially for the current argument, that would mean rejecting any sensibility which involves a negative attitude towards tormenting cats but lacks a negative attitude towards getting your little brother to torment cats.

This enables Blackburn to develop a new reading of **1**, which sees it as the expression of an attitude towards an ethical sensibility. To make things clear he introduces a language that makes value judgements explicit, in which we say “Boo! (x)” or “B! (x)” to express a negative attitude towards x , and “Hooray! (x)” or “H! (x)” to express a positive attitude towards x . In addition we need to be able to express the endorsement or rejection of pairs of attitudes, and he uses a semi-colon to talk about a combination of attitudes or beliefs. So we could say “H! ([B!(x); H!(y)])” to express a positive attitude towards any ethical sensibility which involves negative attitudes towards x and positive attitudes towards y .

If we understand **1** to be the expression of a negative attitude towards a sensibility which involves the conjunction of having a negative attitude towards tormenting cats, but lacks a negative attitude towards getting your brother to torment cats then we can express it as follows:

1''. Boo! ([Boo!(Tormenting the cat)]; (\neg [Boo!(Getting your little brother to torment the cat)])).⁹

In this form there no danger of equivocation between the second premise and the antecedent of the first because the contribution of “tormenting the cat is bad” to the meaning of the conditional **1** is now given in terms of the sentiment that is expressed in **2**.

5.2.2 Earning truth

So, Blackburn has found a way the expressivist can make sense of moral arguments. His next step is to attempt to earn the right to the notion of moral truth:

In effect, quasi-realism is trying to earn our right to talk of moral truth, while recognizing fully the subjective sources of our judgements, inside our own attitudes, needs, desires, and natures. The sense of subjectivity triggers all kinds of wild reactions. Can the projectivist take such things as obligations, duties, the “stern daughter of the voice of God” seriously? How can he if he denies that these represent external, independent, authoritative requirements? (Blackburn, 1984b, p. 197).

⁹In *Spreading the Word* Blackburn’s formulation of this attitude doesn’t negate Boo!(Getting your little brother to torment the cat). He seems to use “A;B” to mean something like the conditional “A→B”. However, Hale (1986, pp. 73–74) points out that the very problem which Blackburn is attempting to solve here is to “*explain* how we come to speak conditionally of our commitments” (Hale, 1986, p. 74), and so it would be problematic to introduce conditionals at this stage without any further explanation. However, as Hale notes, Blackburn can get everything he needs by reading “A;B” as referring to the conjunction “A&B”.

We saw in his response to the Frege-Geach problem Blackburn's introduction of the idea of an ethical sensibility, which he defines as an input-output function, taking representations of the world as its input, and producing moral attitudes as its output. He argued that ethical sensibilities which involve conflicting attitudes (such as a sensibility which has a negative attitude towards tormenting cats but lacks a negative attitude towards getting your little brother to torment cats) should be rejected. He argues that this gives us "a right to a notion of an *improved* set of attitudes [because it] give us some right to a notion of the *coherence* and *consistency* of such a set" (Blackburn, 1984b, p. 197). And we can use this to make sense of the idea of a "best possible set of attitudes" and with that earn our right to the notion of truth:

The simplest solution is that we define a 'best possible set of attitudes', thought of as the limiting set which would result from taking all possible opportunities for improvement of attitude. Saying that an evaluative remark is true would be saying that it is a member of such a set, or is implied by such a set. Call this the set M^* . Then if m is a particular commitment, expressing an attitude m :

m is true = m is a member of M^* (Blackburn, 1984b, p. 198).¹⁰

According to Blackburn this means that if *Murder is wrong* is part of the ideal set of moral attitudes then it will genuinely be true that murder is wrong. A negative attitude towards murder is an attitude which a good person should have, and so we should judge the sentence "murder is wrong" to be true because of this. What it is for a moral sentence to be true is simply being part of the ideal set of moral attitudes.

At this point, it may seem that Blackburn's position is little different from McDowell's. Both defend making realist claims by showing that we can rationally justify preferring one ethical sensibility over others. The key thing to remember here is the difference in their starting points. Using the arguments we saw in §5.1.1 ("Blackburn's arguments for non-cognitivist projectivism") Blackburn argued that we should start off by viewing moral judgements as primarily expressing desires rather than beliefs, and that we should view our moral experience as a result of our projecting moral properties onto the world, rather than detecting them. And it is from that point that he attempts to earn the right to truth using arguments similar to McDowell's. But McDowell does not think we can do so from a starting point which doesn't include genuine moral facts and our ability to detect them (he attempts to justify this claim using the contaminated response and

¹⁰Blackburn notes that this is just a sketch of the simplest projectivist account of truth. He doesn't endorse this view, he is simply using it to give a rough idea of what a projectivist account of truth would be like (Blackburn, 1984b, p. 198, fn. 10).

disentangling arguments, which we'll see in §5.4). According to McDowell we should understand successful ethical sensibilities as involving the detection of genuine moral facts; there is no coherent starting point at which it is right to talk about moral properties as merely being projected onto the world, since we can only understand moral experience if we understand what it is for there to be moral properties.

5.2.3 The problem of mind dependence

Blackburn points out that simply being able to hold that first-order moral claims like "Murder is wrong" are true isn't enough to make projectivism consistent with all of "the important surface phenomena of ethics" (Blackburn, 1988, p. 362). Instead the projectivist must also be able endorse sentences which seem to make metaethical claims:

on a particular moral issue ... one does not just express a desire that the thing should not happen, but one does so while feeling that one's desires on such a matter are *right*. (Blackburn, 1985/1993b, pp. 152–153).

The quasi-realist needs to find a way to hold true sentences such as the following, despite their apparent conflict with projectivism:

- Murder would be wrong even if I didn't think it was.
- The fact that murder is wrong is independent of me, my responses, and my sentiments.
- "even if we had approved of it or enjoyed it or desired to do it, bear-baiting would still have been wrong" (Blackburn, 1985/1993b, p. 153).

Blackburn's response to such sentences is precisely the same as that for first-order moral sentences, he argues that "Since these ethical opinions are unattractive, they must be judged incorrect" (Blackburn, 1988, p. 368). He argues that someone who thought the wrongness of murder depended only upon their own views would be a worse person than someone who thought murder's wrongness was independent of them, and so it is morally right to endorse these sentences in order to commend the attitudes they express. A set of moral attitudes which included the attitude that murder is only wrong because I think it is would be morally inferior to one which didn't include this attitude, and so should be rejected. And so, "The quasi-realist will see [such sentences] as a proper, necessary expression of an attitude toward our own attitudes." (Blackburn, 1985/1993b, p. 153).

For example, Blackburn's response to the last of the three above examples goes as follows:

Perhaps the nicest example comes from counterfactuals that seem to assert an anti-projectivist mind-independence of moral facts: 'even if we had approved

of it or enjoyed it or desired to do it, bear-baiting would still have been wrong' can sound like a second-order realist commitment directly in opposition to projectivism. But in fact, on the construal of indirect contexts that I offer, it comes out as a perfectly sensible first-order commitment to the effect that it is not our enjoyments or approvals to which you should look in discovering whether bear-baiting is wrong (it is at least mainly the effect on the bear). (Blackburn, 1985/1993b, p. 153).

Blackburn's argument here goes as follows: the sentence 'even if we had approved of it or enjoyed it or desired to do it, bear-baiting would still have been wrong' expresses a sentiment of approval to thinking that bear-baiting's wrongness is independent of our enjoyment and desires. This is the correct opinion for a moral person to hold: it would be endorsed by our first-order moral theory, and would be part of the ideal set of moral sentiments. This means that the sentence is true.

5.2.4 The conflict between quasi-realism and non-cognitivism

In response to the problem of mind-dependence we saw how Blackburn argued that we could hold that sentences like "even if we had approved of it or enjoyed it or desired to do it, bear-baiting would still have been wrong" (Blackburn, 1985/1993b, p. 153) are true. Such sentences should be understood to be the expression of moral sentiments, and so their truth or falsity should be judged on whether they express the appropriate sentiments for a moral person to have. But if it is true that the wrongness of bear-baiting is mind-independent, then what are we to make of the non-cognitivist projectivism with which Blackburn started off? There is at least a *prima facie* conflict between holding that moral facts are not dependent upon us, whilst also holding that the reason things seem to have moral properties is because we project our moral attitudes onto our picture of the world. Either they are just the projection of our attitudes or they aren't! Blackburn responds to this apparent problem by arguing that it "confuses two different contexts" (Blackburn, 1985/1993b, p. 157), and that once we make clear the different contexts in which we should understand these two claims then the conflict disappears.

Blackburn argues that the non-cognitivist projectivist aspect of his position is an *explanation* of what we are doing when making moral claims, and that this doesn't conflict with quasi-realism:

Remember here that a projectivist who avails himself of quasi-realism can assert those tantalizing expressions of apparent mind-independence: it is not my sentiments that make bear-baiting wrong; it is not because we disapprove of it that mindless violence is abominable; it is preferable that the world should be a beautiful place even after all consciousness of it ceases. The explanation of what we are doing when we say such things in no way impugns our right to hold them, nor the passion with which we should do so. (Blackburn, 1985/1993b, p. 157).

And a little later he adds:

When the context of discussion is that of first-order commitment, [the quasi-realist] is as solid as the most virtuous moralist. It is just that the explanation of why there are obligations and the rest is not quite that of untutored common sense. It deserves to be called anti-realist because it avoids the view that when we moralize we respond to, and describe, an independent aspect of reality. (Blackburn, 1985/1993b, p. 157).

Here Blackburn seems to be distinguishing between two different contexts in which we're asking two different questions. When discussing first-order commitments about whether our moral judgements are merely mind-dependent we can truthfully assert that they are not. But that doesn't mean that when *explaining* moral talk we can't endorse the projectivist line and describe moral talk as a response to a reality which contains no moral facts.

But so far this still all seems simply inconsistent. Why doesn't the non-cognitivist projectivist explanation of moral talk undermine our first-order claims about the mind-independence of moral facts? Blackburn's response is to claim that there are limits to the contexts in which questions about the dependency of moral facts make sense:

If one attempts to discuss external questions, one must use a different approach — in my case, a naturalism that places the activities of ethics in the realm of adjusting, improving, weighing, and rejecting different sentiments or attitudes. The projectivist, then, has a perfect right to confine external questions of dependency to domains where real states of affairs, with their causal relations, are in question. The only things in this world are the attitudes of people, and those, of course, are trivially and harmlessly mind dependent. But the projectivist can hear no literal sense in saying that moral properties are made for or by sensibilities. They are not in a world where things are made or unmade — not in this world at all, and it is only because of this that naturalism remains true. (Blackburn, 1988, p. 368).

Blackburn distinguishes between internal and external dependency questions. In the internal context we should say that moral facts are mind-independent because that is what the good person would say, and what we would say when straightforwardly immersed in moral talk — the claim that moral facts are mind-independent is an expression of the attitude that a good person would have. In contrast, external dependency questions don't take first-order claims at face value, but instead look to things like the causal relations which lead to our moral talk. In the above passage Blackburn claims that we should confine external questions of dependency to contexts where we are talking about beliefs that are actually caused by real states of affairs. There are no moral facts out there in the world causing our moral beliefs, so the external question about moral facts doesn't make sense. This means that when we consider questions about the mind-dependence

of moral facts only the internal context is available, and in that context moral facts are mind independent; so we can safely hold that moral facts are mind independent.

This explains why Blackburn makes claims such as the following:

- “Talk of dependency is moral talk or nothing.” (Blackburn, 1988, p. 368).
- “[The quasi-realist] affirms *all that could ever properly be meant* by saying that there are real obligations” (Blackburn, 1985/1993b, p. 157, emphasis in original).
- “[The quasi-realist denies] that moral truth is mind dependent in the only sense possible” (Blackburn, 1988, p. 368).

So, there is only one way of reading questions about the mind-dependency of moral facts; we should understand such question as internal to the moral domain, in which holding that moral facts are mind-independent is the right thing to say and is true.

But what’s Blackburn’s justification for claiming that the external reading of moral dependency questions doesn’t make sense? In the longer passage quoted above Blackburn seems to simply assert that in the external context such questions don’t make sense because in the moral domain we are not talking about real states of affairs and their causal relations to our beliefs. But why does that prevent the external reading from making sense? The following passage is illuminating:

there is only one proper way to take the question “On what does the wrongness of wanton cruelty depend?”: as a moral question, with an answer in which no mention of our actual responses properly figures. There *would* be an external reading if realism were true. But antirealism acknowledges no such state of affairs, and no such issue of dependency. (Blackburn, 1988, p. 367, emphasis in original).

Here Blackburn acknowledges that the external reading of the dependency question would make sense if realism were true, but he thinks realism is false and so it doesn’t make sense.

But this acknowledgement, that “There *would* be an external reading if realism were true.” undermines Blackburn’s entire strategy. If he gave us reason to think that external questions about the dependency of moral facts were *inherently* incoherent then his position would be fine. But here we have him acknowledging that such questions *could* make sense — there is nothing about the domain of moral talk itself which makes such questions incoherent. Instead he claims that it is simply the fact that realism is false and that there are no real moral properties which makes them inherent. But that’s not how language works! The coherence of a question is not dependent upon whether or not the things talked about exist — it’s perfectly coherent to deny that Father Christmas exists and to hold that a young child’s belief in Father Christmas is entirely dependent upon the things they have been told by their parents. When talking about Father Christmas the external question of whether belief in him is mind-independent makes perfect sense, despite the

fact that he doesn't actually exist. Why not in the case of ethics? It can't be merely the falsity of realism that renders the external question incoherent, instead it must be something about the moral domain itself. And Blackburn has given us no reason to show that such questions are inherently incoherent, and has even acknowledged their possible coherence by saying that they would be coherent if there were genuine moral properties.

However, in this evaluation of Blackburn's quasi-realist move I have only considered his argument as it appears in the papers "How to be an Ethical Antirealist" (Blackburn, 1988) and "Errors and the Phenomenology of Value" (Blackburn, 1985/1993b). Blackburn has written extensively on this problem, but since the focus of this chapter is not on Blackburn himself, but instead on the debate between Blackburn and McDowell, I have not had the opportunity to explore his arguments in sufficient depth to conclude that he is unable to answer to this concern. But at the very least it seems fair to say that he does need to say more than he does in those two papers to explain why the external questions should be ignored when assessing second-order moral claims.

5.3 Explanatory tests

We have seen two different views of how moral claims can be true: Blackburn's projectivism and McDowell's no-priority view. According to Blackburn we can explain moral practice by starting off with a view of reality which contains no moral facts, and see our moral experience as a result of our projecting our moral sentiments onto reality. But he argues that from this starting point we can still earn our right to the notion of moral truth, by saying that a moral judgement is true if it expresses a sentiment that we should endorse.

In contrast, McDowell's view is (at least on the surface) much more straightforward. There are moral facts and they make our moral judgements true or false. We can't understand these facts independently of our understanding of our moral sentiments, but that isn't problematic for moral realism because the same is true for secondary quality facts (such as facts about colours). We also can't understand our moral sentiments and experience without reference to apparent moral facts, and so any explanation of our moral sentiments and experience must make space for at least the possibility of there being moral truths. Moral talk is able to earn the right to the notion of truth by its own lights — we should hold that there are moral truths if moral reasoning gives us a boat that floats.

5.3.1 The significance of the quasi-realist starting point

Blackburn argues that his position is superior to realist positions like McDowell's because he can earn truth without simply assuming that there are moral facts, and working from a starting point which involves no moral features:

I distinguish then between the ingredients with which you start, and what you can legitimately end up saying as you finish. To place ethics, I deny that we can help ourselves to moral features and explanations from the beginning. We have to see them as constructions, or as I call it, projections, regarding ourselves in the first instance as devices sensitive only to natural facts, and producing only explicable reactions to them. The aim is to explain and make legitimate the emergence of full-blown ethics on this austere basis. But there is no need to deny, as the error theorists did, that the full blown system is in order as it is. Nor is there any need to regret apparently realistic features of it, if these can be earned from the slender basis. (Blackburn, 1991, p. 16).

Unlike McDowell, Blackburn starts off with a view of the natural world which contains no moral features and instead sees them as projections onto a value-free reality. But, as we've seen above, from this basis he goes on to attempt to earn our right to the notion of truth. This is the quasi-realist project. In contrast, McDowell's view starts off by taking our moral talk at face value, and with a view of the world which involves value facts. And so, Blackburn's argues, McDowell fails to fully earn the right to talk of moral truth, because he simply assumes there are value facts from the beginning.

In "Realism, Quasi, or Queasy?" Blackburn (1993d) explains his position in the context of the dispute between imminent and external realism. According to imminent (or "internal") realist positions, such as McDowell's moral realism, to earn our right to be realists about a discourse all we need do is work from within using the tools of that discourse. If this is correct then to justify realism for a discourse all we need to do is show things like that truth is unique and that the truth and our actual opinions can come apart (Blackburn, 1993d, p. 370). In "Projection and Truth in Ethics" McDowell adds that to earn our right to the notion of truth we must be able to show that there can be good reasons for changing our minds, i.e. there must be scope for "rationally induced improvements" (McDowell, 1987/1998l, p. 156), instead of "merely manipulative persuasion" (McDowell, 1987/1998l, p. 156). This is the kind of work we saw McDowell doing towards the end of the last chapter in order to respond to Mackie's arguments against moral realism. If McDowell's arguments there seemed unsatisfying that is because they only attempted to be internal arguments for realism, and didn't involve responses to external questions.

In contrast, external realists like Blackburn think that "there is a further external, metaphysical issue over whether the right theory of [an] area is realistic" (Blackburn, 1993d, p. 368). When judging whether we should believe in the reality of the facts involved in a discourse we need to not only test the discourse by its own lights, but must also look beyond our first-order theory and consult "evidence from its surroundings, or in other words from other things we think, for example about explanation, truth, fact, reality, and our knowledge of it" (Blackburn, 1993d, p. 371).

As we saw in §5.1.1 (“Blackburn’s arguments for non-cognitivist projectivism”), Blackburn’s conclusion in response to these external questions is that we should start off as anti-realists about morality — unlike McDowell we shouldn’t simply presume that we understand moral facts and can have knowledge of them — because moral facts do not fit into our overall scientific world view of what facts there can be and how we can know about them. The apparent moral facts are merely projected onto the world, leading us to have experience of them without the need for any genuine moral properties. But Blackburn argues that from this projectivist starting point he can still earn the right to the notion of truth in ethics. He argues that McDowell doesn’t do any of this kind of work, but instead ignores the external questions and simply takes our moral talk at face value without sufficiently justifying doing so. And so, Blackburn concludes, only the quasi-realist is justified in claiming to have earned the right to the notion of truth in ethics, since only the quasi-realist has a satisfactory explanation of what makes moral judgements true.

5.3.2 The two different explanatory tests

In response to this criticism McDowell argues that Blackburn has focused on the wrong explanatory test:

The right explanatory test is not whether something pulls its own weight in the favoured explanation . . . , but whether the explainer can consistently deny its reality. (McDowell, 1985/1998m, p. 142).

In the case of values this gives us the following two explanatory tests:

Explanatory test 1 Whether the explainer can consistently deny the reality of values.

Explanatory test 2 Whether value facts pull their weight in an explanation.

Test **2** is the kind of thing the external realist looks for, for example, external anti-realists like Blackburn argue that the supposed value facts do no work in explaining why we have the moral beliefs we do. Contrast our awareness of facts about values with our visual experience of the world; in the latter case we can tell intricate stories about sense perception and the physical processes involved. These stories fit into our general understanding of reality and how we are able to know about facts, and the facts, such as the fact that something is square, play an essential role in such an explanation of how we know them. In the moral case we can tell no naturalised story explaining how the value facts lead us to have the moral beliefs we do. Value facts fail the second of the two explanatory tests McDowell identifies. And so, Blackburn concludes that we should be anti-realists about values.¹¹

¹¹As with much of Blackburn’s position, his response to such questions is subtle. As we saw in the last section, in “Realism, Quasi, or Queasy?” he responds to the “external, metaphysical issue over whether

But, as we saw in the above quotation, McDowell claims that this is not the explanatory test we should be focusing on, and that we should instead focus on test **1**, which is the question of whether our explanation of value judgements and value experience is consistent with denying the reality of values. He argues that Blackburn's non-cognitivist projectivist explanation of moral judgements fails because the explanation is incompatible with a world which includes no values, because Blackburn's explanation of moral sentiments requires reference to moral properties. If this is the case then that means that we must be committed to there being values, right from the start; Blackburn's austere starting point is incoherent and the quasi-realist project is unnecessary. I shall go into McDowell's argument for this in the next section, but first it's important to note, and respond to, a potential worry about Blackburn's quasi-realism.

As we saw above, Blackburn thinks that our explanation of moral judgements shouldn't start off with the presumption that there are genuine moral properties, since such properties fail the required external tests for realism. But he thinks that from that projectivist starting point we can earn our right to the notion of truth in ethics. If this is the case then why would it matter that his explanation is incompatible with the denial of value facts?

McDowell's aim, in the two arguments I shall look at next, is to show that non-cognitivism is incompatible with projectivism. Blackburn's starting point aims to explain

the right theory is realistic" (Blackburn, 1993d, p. 368) by concluding that for values it is not. This is because ethical sensibilities have no "naturalized dependency" (Blackburn, 1993d, p. 377) on ethical properties. However, in "Just Causes" Blackburn (1991, pp. 11–16) suggests a way we can understand moral properties to be "causally relevant, in spite of having a projected origin" (Blackburn, 1991, p. 13). He refers to Jackson and Pettit's work on "program explanation" (Jackson & Pettit, 1990) which is the view that supervening properties can be causally relevant, despite not being causally efficacious, because "[their] realization programs for the realization of a lower-order efficacious property" (Jackson & Pettit, 1990, p. 115). For example, when explaining why a glass broke it is its molecular structure which is causally efficacious. Fragility is a supervening property, which programs for the type of molecular structure which is likely to break easily. So the fact that glass is fragile ensures that it will have the kind of molecular structure that is likely to break easily. And so,

the higher-order, inefficacious property is causally relevant to the event produced, because its realization programs for the realization of a lower-order efficacious property and, in the circumstances, for the occurrence of the event in question. (Jackson & Pettit, 1990, pp. 114–115).

Blackburn suggests that the quasi-realist might be able to make the same move in the moral case — it is correct to say that injustice caused the revolution despite the fact that injustice is not a natural property and not causally efficacious. This suggests a way the quasi-realist can respond to the external realist question with a negative answer (and so initially reject the reality of moral properties) but still conclude that there is nothing wrong with our talk which involves making use of moral properties in our explanations, even causes, of events. (I should note, however, that Blackburn is very tentative in his endorsement of this approach. And also, Jackson and Pettit's view has many subtleties which I lack the space to go into here.)

our moral judgements as being expressions of attitudes rather than expressions of beliefs. Ultimately Blackburn concludes that moral judgements are true if they express the right attitudes, but he still needs his starting point to be consistent for it to be the foundation from which he attempts to earn truth. McDowell aims to show that you cannot explain moral judgements as being mere expressions of attitudes, and that no explanation of moral judgements can be given on that basis. We can't see moral properties as being mere projections and must regard ourselves as sensitive to genuine moral properties to explain our reactions to them. The austere basis from which Blackburn aims to start fails to be a coherent account of moral sentiments. If McDowell is correct then the only explanation of our moral judgements involves genuine moral properties right from the start — we cannot begin to explain moral judgements without presuming that there are moral properties. And if we accept that there are moral properties right from the start then there is no need for the quasi-realist's project, since we can simply say that it is these facts about values which make our moral statements true or false, and not whether or not they align with the ideal set of moral sentiments.

So, we can conclude that McDowell is right to stress the significance of the latter explanatory test, but that alone doesn't show that the other explanatory test isn't also worth considering. Of course, McDowell is right to argue that any explanation which denies certain facts must be consistent with denying the reality of these facts, but shouldn't we also be worried about whether value facts pull any explanatory weight? If we cannot explain how the value facts pull their weight then perhaps we should reject their reality? Or, at the very least, we should attempt to develop an explanation of how value facts are able to pull their weight; as Blackburn states "there can never be an *a priori* right to claim that our activity in making judgements *X* permits of no explanation . . . You just have to try the various explanations out." (Blackburn, 1985/1993b, p. 163). In §5.5 I shall look at how McDowell might respond to this explanatory test, but first I shall look at McDowell's argument that Blackburn fails the first.

5.4 The first explanatory test: McDowell's contaminated response and disentangling arguments

McDowell develops two closely related arguments which attempt to show that Blackburn fails this explanatory test, both focusing on the combination of non-cognitivism with projectivism. The contaminated response argument attempts to show that the non-cognitivist cannot give an account of the sentiments which are apparently projected onto the world when we make a moral judgement. The disentangling argument attempts to show that the non-cognitivist cannot give an account of the aspects of the world which our moral judgements are responses to. If successful, either of these attacks would be

fatal for Blackburn since they attempt to show that he cannot adequately explain moral judgements whilst denying the reality of values, and so his austere starting point is not an available position. McDowell argues that only the cognitivist can give a satisfying account of moral judgements.

I shall conclude that neither of these arguments is likely to be fatal for the non-cognitivist, since both rely on premises which it seems Blackburn could simply reject. In the case of the disentangling argument the premise is highly controversial, although it at least seems that we might end up concluding it is true. So when that work is done the argument may undermine Blackburn's position, but as it stands alone the argument is ineffective. And in the case of the contaminated response argument I shall suggest that McDowell could run a similar argument with a less questionable premise, but it is still a premise which he has failed to provide sufficient argument for if he wants to claim to have shown Blackburn's starting point to be incoherent.

5.4.1 The disentangling argument

The disentangling argument was developed by McDowell in the papers "Virtue and Reason" (McDowell, 1979/1998n) and "Non-cognitivism and Rule-following" (McDowell, 1981/1998i). It attempts to show that Blackburn's non-cognitivism prevents him from giving an adequate explanation of moral judgements, and that only the cognitivist can give a satisfactory account of the things in the world which lead us to make moral judgements.

According to the non-cognitivist projectivism that forms Blackburn's starting point when we see the world to have certain properties we will have a certain non-cognitive attitude — a moral sentiment — and project that out onto our experience of the world. Moral judgements are expressions of these sentiments. McDowell argues that this view commits Blackburn to being able to "disentangle" the items in the world which moral judgements are responses to from the moral sentiments:

Typically, non-cognitivists hold that when we feel impelled to ascribe value to something, what is actually happening can be disentangled into two components. Competence with an evaluative concept involves, first, a sensitivity to an aspect of the world as it really is (as it is independently of value experience), and, second, a propensity to a certain attitude — a non-cognitive state that constitutes the special perspective from which items in the world seem to be endowed with the value in question. . . . in making value judgements [we] register the presence in objects of some property they authentically have, but enrich [our] conception of this property with the reflection of an attitude. (McDowell, 1981/1998i, pp. 200–201).

The disentangling requirement is the requirement that the non-cognitivist must be able to separate the first of these components from the second — to explain their position they must be able to make sense of our "sensitivity to an aspect of the world as it really is (as

it is independently of value experience)” independently from their understanding of our propensity to moral attitudes. As we shall see this is ultimately the challenge to identify the things in the world which a competent moraliser will judge to be good or bad (the things that our moral sentiments are responses to), and to do so without presuming an understanding of the moral sentiments themselves. In summary:

Disentangling requirement: A non-cognitivist account of moral judgements must be able to identify the items in the world that moral judgements are responses to, and it must do so without making use of our moral sentiments themselves — it can’t identify these items in the world simply as those which will elicit moral attitudes.

But why is the non-cognitivist projectivist committed to being able to identify the items in the world which our moral sentiments are responses to? In his exegesis of this argument Miller (2013, pp. 240–252) distinguishes between two reasons for why the non-cognitivist must be able to disentangle the relevant items in the world from our moral attitudes.

The first reason why the non-cognitivist must be able to disentangle the items in the world from our sentiments is because non-cognitivists are committed to thinking that moralising — our uses of evaluative language — is not just “sounding off” (McDowell, 1981/1998i, p. 217), but a genuine conceptual process. When we make moral judgements we are “going on in the same way” (McDowell, 1981/1998i, p. 216) in that moral judgements have something in common that distinguishes them from non-moral judgements. A competent moraliser does not express sentiments towards items in the world randomly, but does so “going on in the same way”. And in order for this to be the case the items to which their moral sentiments are directed must fall into groups that share natural features — there must be something in common between the items in the world that their moral judgements respond to which makes moral judgements part of the same practice. As Miller puts it:

If we cannot identify a genuine aspect of the world as that to which the non-cognitive element in moral judgement is sensitive, it is doubtful whether we can view the repeated tokenings of evaluative language as instances of ‘going on in the same way’. All we would have would be expressions of non-cognitive sentiment directed at a heterogeneous collection of items: not enough to ground a conceptual practice. (Miller, 2013, p. 244).

So, for moral judgements to be part of a genuine conceptual practice, and not just random “sounding off” (McDowell, 1981/1998i, p. 217), we must be able to identify something which the items in the world that our moral judgements are responses to have in common. For the cognitivist about values, the answer to this question is straightforward: the items in the world which the competent moraliser responds to when they make moral judgements have in common the fact that they have moral properties. What is it that the items in

the world which competent moralisers judge to be good have in common? The fact that they have the property of being good.

Clearly the non-cognitivist can't say this, since they deny that we should understand our moral judgements as responses to genuine moral properties such as goodness and badness. Of course there is something which all items in the world which elicit moral judgement have in common in the non-cognitivist projectivist account, and that is that these items all elicit moral sentiments in competent moralisers. But McDowell argues that the non-cognitivist can't identify the items in the world that moral judgements are responses to as those which will elicit moral attitudes — those that the competent moraliser will judge to be good or bad — because “[the tendency to elicit the moral attitude] is not a property to which the attitude can coherently be seen as a response” (McDowell, 1981/1998i, p. 217).

The second justification for the disentangling requirement that Miller identifies is related to Blackburn's use of the notion of an ethical sensibility. As we saw in §5.2, Blackburn's responses to the Frege-Geach problem and the problem of mind dependency relied on the claim that we express moral attitudes not only towards acts and items in the world but also towards ethical sensibilities. This is problematic because ethical sensibilities are input-output functions:

We can usefully compare the ethical agent to a device whose function is to take certain inputs and deliver certain outputs. The *input* to the system is a representation, for instance of an action, or a situation, or a character, as being of a certain type, as having certain properties. The *output*, we are saying, is a certain attitude, or a pressure on attitudes, or a favouring of policies, choices, and actions. Such a device is a function from input to output: an ethical sensibility. (Blackburn, 1998, p. 5).

If the disentangling requirement can't be met and the only way to identify the input — the items in the world to which our moral judgements are responses — is as those that will elicit the output — moral attitudes — then there will be no way to distinguish between the input and output in an ethical sensibility. But if you can't distinguish between the input and the output in an ethical sensibility then it simply doesn't make sense to talk of ethical sensibilities at all. To define an ethical sensibility you need to be able to independently identify the items in the world that the moral sentiments are responses to, you can't simply define them as those which will elicit the appropriate output. And so Blackburn is committed to being able to satisfy the disentangling requirement.

What this means is that the non-cognitivist needs to be able to provide something along the lines of:

x will lead S to have the subjective response of judging x to be good if and only if x is N .

Where S is a competent moraliser, and N is a natural property that can be identified independently of our moral attitudes and subjective responses.

The uncodifiability of morality

Before going further it's important to note that McDowell's no-priority view, which I introduced in the last chapter, cannot be used to demonstrate that the disentangling requirement can't be met. At first sight it may seem that if the second half of the no-priority view — the claim that *understanding the property of goodness requires understanding of moral responses* — were correct then it would be impossible to satisfy the disentangling requirement, but this is wrong. If this half of the no-priority view of moral concepts were correct then there would be no way to understand the property of goodness — no way to grasp the sense of 'good' — if you didn't understand moral sentiments. This is not relevant here because the disentangling requirement is at the level of *reference* rather than sense. To satisfy the disentangling requirement all that needs to be done is the *identification* of the items in the world which moral judgements are responses to — the identification of the things a competent moraliser would call 'good' — without making use of your understanding of the moral sentiments. There is nothing in the no-priority view to show this cannot be done. Instead, the no-priority view makes a claim about the limits on our ability to grasp the sense of 'good', not on our ability to know its extension. The no-priority view has nothing to say about the possibility that there is a natural predicate 'N' that has a different sense from 'good', but which is nevertheless co-extensional with 'good'. If there were such a predicate then it would be possible to identify the items which the competent moraliser will respond to with moral sentiments, without using our concept *good* and without relying on our understanding of the sentiments themselves. This means that McDowell needs something much more substantial than the no-priority view to demonstrate that the disentangling requirement can't be met.¹²

One way of filling in the right-hand side of the above biconditional would be provided if we endorsed a simplistic utilitarianism and held that something is good if and only if it maximises happiness. This would give us the following biconditional:

¹²Miller's exegesis of McDowell accuses him of failing to recognise this fact (Miller, 2013, pp. 240–252). According to Miller's reading McDowell argues only that the non-cognitivist can't grasp the sense of 'good' and then jumps to the conclusion that this means they can't know the extension of 'good' — according to Miller McDowell simply assumes that there is no alternative predicate 'N' which has a different sense, but the same extension as 'good'. In contrast, according to my reading, McDowell does provide an argument that there is no such alternative predicate 'N', and although his argument relies on a questionable assumption, he isn't guilty of the confusion between sense and reference of which Miller accuses him.

x will lead S to have the subjective response of judging x to be good if and only if x maximises happiness.

If this simplistic form of utilitarianism were true then the non-cognitivist would have the answer they need to respond to the disentangling challenge. They would be able to identify the items of the world that moral judgements are responses to — they are those involved in the actions which maximise happiness — and do so without having to rely on our moral sentiments.

The problem is, of course, that this simplistic utilitarianism is generally held to be false. Perhaps it's possible for the non-cognitivist to do better and provide an account of what items in the world would be good which doesn't rely on the concept of good, but McDowell claims that is implausible:

this picture fits only if the virtuous person's views about how, in general, one should behave are susceptible of codification . . . But to an unprejudiced eye it should seem quite implausible that any reasonably adult moral outlook admits of any such codification. As Aristotle consistently says, the best generalizations about how one should behave hold only for the most part. If one attempted to reduce one's conception of what virtue requires to a set of rules, then, however subtle and thoughtful one was in drawing up the code, cases would inevitably turn up in which a mechanical application of the rules would strike one as wrong — and not necessarily because one had changed one's mind; rather, one's mind on the matter was not susceptible of capture in any universal formula. (McDowell, 1979/1998n, pp. 57–58).

The key part of this quotation is the claim that “the best generalizations about how one should behave hold only for the most part” (McDowell, 1979/1998n, p. 58); the claim that “how one should live is not codifiable” (McDowell, 1979/1998n, p. 65). And in the later paper “Non-cognitivism and Rule-following” McDowell describes this claim as the rejection of “the assumption that a moral stance can be captured in a set of externally formulable principles — principles such that there could in principle be a mechanical (non-comprehending) application of them that would duplicate the actions of someone who puts the moral stance into practice” (McDowell, 1981/1998i, pp. 214–215), and states that “This assumption strikes me as merely fantastic.” (McDowell, 1981/1998i, p. 215).

In “Virtue and Reason” McDowell (1979/1998n, p. 58, fn. 11) also references Aristotle's formulation of this claim, which is as follows:

goods also give rise to a similar fluctuation because they bring harm to many people; for before now men have been undone by reason of their wealth, and others by reason of their courage. We must be content, then, in speaking of such subjects and with such premises to indicate the truth roughly and in outline, and in speaking about things which are only for the most part true and with premises of the same kind to reach conclusions that are no better. (Aristotle, 1925, 1094b17–23).

Later in “Virtue and Reason” McDowell’s makes a parallel claim in epistemological form: “Occasion by occasion, one knows what to do, if one does, not by applying universal principles but by being a certain kind of person.” (McDowell, 1979/1998n, p. 73), which seems particularly hard to deny. Even though we regularly use universal principles in our moral reasoning we often find that they fail us in hard cases and we instead conclude that in some cases the right thing to do is distinct from that which a universal principle recommends. It seems that in our moral reasoning general principles play the role of heuristics for guiding our thought, rather than as guaranteed tests that are expected to work in all cases.

So, McDowell concludes that “how one should live is not codifiable” (McDowell, 1979/1998n, p. 65). If this is true then the only way to fill in the above biconditional is as follows:

x will lead S to have the subjective response of judging x to be good if and only if x is good.

Which of course is precisely what non-cognitivists like Blackburn can’t do, since it involves good being a genuine property of the world, which, at least at this point in his quasi-realist account, he wants to deny. Blackburn’s austere starting point is intended to explain our moral judgements in terms of sentiments which we project onto the world, and he wants to go on to earn the notion of truth in ethics from there. But if from the very start he must refer to moral properties then he wouldn’t be a quasi-realist at all, but simply a realist.

McDowell’s failure to justify the uncodifiability thesis

So, McDowell’s strategy, both in “Virtue and Reason” (McDowell, 1979/1998n) and “Non-cognitivism and Rule-following” (McDowell, 1981/1998i), is to assert the intuitive force of the claim that morality is not codifiable into rules graspable by someone who’s not a participant in moral practice and then to demonstrate that this claim is incompatible with non-cognitivism. However, the problem with this is that McDowell has not provided us with sufficient argument for the claim that morality is uncodifiable.

McDowell’s justification for the uncodifiability of morality seems to take two forms:

- It is implausible that morality is codifiable.
- We don’t reason about morality using universal principles; all general principles we use in our moral reasoning are expected to hold only for the most part.¹³

¹³In *An Introduction to Contemporary Metaethics* Miller (2013, pp. 240–252) claims that McDowell attempts to use Wittgenstein’s rule-following arguments to justify the uncodifiability thesis. Miller responds to McDowell by pointing out that the rule-following arguments can’t be used to justify the uncodifiability

The first of these reasons is simply an assertion, not an argument. Even more problematic for McDowell, it is a highly controversial one. We saw one attempt to codify morality above: the over-simplified utilitarian view that x is good if and only if x maximises happiness. Clearly this attempt to codify morality fails. But that does not show that a better codification can't be provided, and indeed finding such a codification is essentially the project of all consequentialist theories of ethics, and perhaps all deontological theories too. Consider all the work that has been done to attempt to refine the simplistic utilitarian view and find such a universal principle. If it was universally accepted that morality is uncodifiable then it's hard to believe that this work would have been done.

The second reason looks more hopeful, but also fails to do the work needed to justify the uncodifiability thesis. The problem is that it makes a claim about how we do our moral reasoning, and uses that to justify the much stronger claim, effectively about what moral reasoning could ever be possible. It may well be true that when reasoning about morality we never expect any general principles to apply universally, and expect that our general principles will always have counterexamples, but that doesn't show that there *can't* be any successful universal principles. All it shows is that we don't currently have any such universal principles, that we don't expect to have any, or perhaps that the failing to discover such principles is down to us rather than the principles themselves.

In the later paper "Values and Secondary Qualities" McDowell (1985/1998m) seems to acknowledge the limitations in the disentangling argument. There he gives the following compressed version of the argument:

thesis since the rule-following arguments only demonstrate a limitation on who can grasp the sense of 'good', not on whether there can be any other terms with the same extension as 'good', which is all that is required for the uncodifiability thesis to be false.

However, I think that Miller's reading of McDowell's use of the rule-following arguments here is wrong. McDowell is not using rule-following in an attempt to demonstrate that the uncodifiability thesis is true, instead he's merely using it to undermine "A deep-rooted prejudice about rationality [that] blocks ready acceptance of this [the uncodifiability thesis]." (McDowell, 1979/1998n, p. 58). Consider the following passage:

[Non-cognitivists] may admit that it is often difficult to characterize the authentic property ... that corresponds to an evaluative concept; but they tend to suppose that there *must* be such a thing, even if it cannot be easily pinned down in words. Now there is a profoundly tempting complex of ideas about the relation between thought and reality that would make that "must" seem obvious. But one strand in Wittgenstein's thought about "following a rule" is that the source of the temptation is the desire for a security that would actually be quite illusory. (McDowell, 1981/1998i, p. 203).

In other words, McDowell claims that non-cognitivists "tend to suppose that there *must* be" a way to codify 'good' in no moral terms. This supposition is justified by a "profoundly tempting complex of ideas", but is undermined by Wittgenstein's rule-following arguments.

So, as I read him, McDowell uses rule-following here, not to prove that the uncodifiability thesis is true, but to block a potential reason for thinking that the uncodifiability thesis must be false. However, I don't want to get too bogged down in evaluating competing interpretations of McDowell since it isn't particularly important whose reading is correct. Even if Miller is right, and McDowell is attempting to use rule-following to justify the uncodifiability thesis, that would not be significant, since (as Miller points out), the rule-following arguments cannot be used to prove the truth of uncodifiability.

In the projectivist picture, having one's ethical or aesthetic responses rationally suited to their objects would be a matter of having the relevant processing mechanism functioning acceptably. . . . it pictures the mechanism as something that one can contemplate as an object in itself . . . How, then, are we to understand this pictured availability of the processing mechanism as an object for contemplation, separated off from the world of value? Is there any alternative to thinking of it as capable of being captured, at least in theory, by a set of principles for superimposing values on to a value-free reality? The upshot is that the search for an evaluative outlook one can endorse as rational becomes, virtually irresistibly, a search for such a set of principles: a search for a *theory* of beauty or goodness. One comes to count "intuitions" as respectable only in so far as they can be validated by an approximation to that idea. . . . I have a hunch that such efforts are misguided; not that we should rest content with an "anything goes" irrationalism, but that we need a conception of rationality in evaluation that will cohere with the possibility that particular cases may stubbornly resist capture in any general net. . . . I take it that my hunch poses a question of moral and aesthetic taste, which — like other questions of taste — should be capable of being argued about. The trouble with projectivism is that it threatens to bypass that argument (McDowell, 1985/1998m, pp. 148–149).

Crucial here is that he says that he has merely a "hunch" that there is no theory of goodness; that there is no set of principles which can be understood separated off from the world of value. What played the role of the central premise in the arguments in "Virtue and Reason" (McDowell, 1979/1998n) and "Non-cognitivism and Rule-following" (McDowell, 1981/1998i) is now merely a hunch.

However, in "Values and Secondary Qualities" he states that this doesn't completely undermine the significance of the disentangling argument, arguing as follows:

I take it that my hunch poses a question of moral . . . taste, which — like other hunches — should be capable of being argued about. The trouble with projectivism is that it threatens to bypass that argument (McDowell, 1985/1998m, pp. 149–150).

The disentangling argument shows that if the uncodifiability thesis is correct then non-cognitivism must be false. McDowell has failed to demonstrate that the uncodifiability thesis is true, and so the disentangling argument fails to conclusively demonstrate the falsity of non-cognitivism. However, the truth of the uncodifiability thesis is a live debate in meta-ethics; McDowell for one believes it to be true and rightly notes that it "should be capable of being argued about". In effect, what the disentangling argument shows is that non-cognitivist accounts of our moral judgements are hostages to fortune with respect to this debate. Whilst there is no good reason to believe in the truth of the uncodifiability thesis, other than the fact that it seems highly plausible to some philosophers, non-cognitivism is fine. But if arguments were to be provided which demonstrated that we

should endorse the uncodifiability thesis then one consequence of that discovery would be the fact that non-cognitivist accounts of moral judgements fail. Since Blackburn's starting point depends upon what is essentially non-cognitivism, this means that his claims to have found a way to defend the truth of our moral judgements which is superior to McDowell's is also a hostage to this debate.

5.4.2 The contaminated response argument

McDowell's contaminated response argument was developed in "Values and Secondary Qualities" (McDowell, 1985/1998m, pp. 141–146) and "Projection and Truth in Ethics" (McDowell, 1987/1998l, pp. 157–162). Like the disentangling argument, the contaminated response argument attempts to show that Blackburn is unable to give an adequate explanation of moral judgements from his non-cognitivist starting point. Whereas the disentangling argument questions Blackburn's ability to give an account of the items in the world to which our moral judgements are responses, the contaminated response argument instead attacks Blackburn's ability to characterise the moral sentiments. In essence, McDowell claims that from the non-cognitivist projectivist starting point there is no way to give a satisfactory account of the sentiments that are allegedly projected onto the world in our moral judgements, since any account of these sentiments requires us to use moral concepts and so commits us to a cognitivist account of morality. (The need to make use of moral concepts would be devastating for Blackburn because moral concepts represent the very moral properties which he wants his starting point to be free of.)

As we saw above, Blackburn attempts to start his quasi-realist project with an explanation of moral judgements and experience which excludes moral properties. According to Blackburn's explanation our seemingly moral experiences are actually results of our having certain sentiments when we experience certain things in the world, and then projecting these sentiments onto the world. From that austere basis he hopes to earn truth, but if that austere picture can't be drawn then his project fails. So, according to Blackburn's starting point, when I see Tommy drowning in a lake I will have a certain positive sentiment towards saving Tommy, which leads to it seeming to me that saving him would be good. The sentiment of a positive attitude towards saving drowning children, when combined with my belief that Tommy is a drowning child, motivates me to act and save Tommy. McDowell argues that this account commits Blackburn to certain explanatory requirements:

The point of the image of projection is to explain certain seeming features of reality as reflections of our subjective responses to a world that really contains no such features. Now this explanatory direction seems to require a corresponding priority, in the order of understanding, between the projected response and the apparent feature: we ought to be able to focus our thought

on the response without needing to exploit the concept of the apparent feature that is supposed to result from projecting the response. (McDowell, 1987/1998l, p. 157).

McDowell points out that since Blackburn's starting point is committed to there being no genuine moral properties in the world that cause us to make our moral judgements, Blackburn needs to be able to give an account of the response (the moral sentiment) which is involved in moral judgements without making use of moral concepts. If he cannot give us such an account of the moral sentiments — one which doesn't require the use of moral concepts like *good* and *bad* as potential properties of the world — then he will have failed to give an account of moral judgements which is compatible with the rejection of moral realism.

On the realist view the most straightforward account of moral experiences and sentiments is that they are responses to our awareness of the moral properties — a particular sentiment is a moral one if and only if it involves belief in the instantiation of a moral property. Blackburn cannot tell such a story because he denies that moral judgements involve the belief that there are moral properties in the world — the story told from his starting point is non-cognitivist. McDowell points out that this means that Blackburn is committed to being able to give an account of the moral sentiments without making use of the concepts of moral properties, but he argues that without presuming an understanding of these concepts Blackburn will be unable to distinguish between moral and non-moral sentiments. And if Blackburn has no way of identifying moral sentiments, then his non-cognitivist projectivist view is doomed.

McDowell attempts to support his claim that Blackburn cannot give an account of moral sentiments by drawing an analogy with the comic, where he claims that comic sentiments also cannot be characterised without relying on the idea of things actually being comic:

What exactly is it that we are to conceive as projected onto the world so as to give rise to our idea that things are funny? "An inclination to laugh" is not a satisfactory answer: projecting an inclination to laugh would not necessarily yield an apparent instance of the comic, since laughter can signal, for instance, embarrassment just as well as amusement. Perhaps the right response cannot be identified except as amusement; and perhaps amusement cannot be understood except as finding something comic. I need not take a view on whether this is correct. But if this is correct, there is a serious question whether we can really *explain* the idea of something's being comic as a *projection* of that response. ... differentiating some exercises of that unspecific propensity as cases of *amusement* is something we have to learn, and if the suggestion is correct, this learning is indistinguishable from coming to find some things comic. Surely it undermines a projective account of a

concept if we cannot home in on the subjective state whose projection is supposed to result in the seeming feature of reality in question without the aid of the concept of that feature, the concept that was to be projectively explained. (McDowell, 1987/1998l, p. 158, emphasis in original).

In this passage McDowell starts by considering one way we could identify comic sentiments: comic sentiments are those that involve an inclination to laugh. But then he points out that this doesn't work since many other sentiments involve an inclination to laugh — laughter can also signal embarrassment as well as amusement. What we need at this point is something along the following lines:

r is a subjective response involved in *S* judging something to be funny if and only if *S* has *r* any time she has an inclination to laugh.

As we've just seen, this way of identifying subjective responses of amusement fails because we can have an inclination to laugh in many other cases, not just those which we find amusing.

McDowell's next step is to suggest that perhaps the only way we can identify which subjective responses are those of amusement is with something like this:

r is a subjective response involved in *S* judging something to be funny if and only if *S* has *r* any time she finds something to be funny.

If this is the only way we have for identifying the relevant subjective responses — if the only way of identifying which subjective states are cases of amusement uses the concept of things actually being amusing — then the non-cognitivist projectivist account of amusement is in serious trouble. As McDowell argues:

it undermines a projective account of a concept if we cannot home in on the subjective state whose projection is supposed to result in the seeming feature of reality in question without the aid of the concept of that feature, the concept that was to be projectively explained. (McDowell, 1987/1998l, p. 158).

The non-cognitivist projectivist about humour seeks to explain the fact that things seem to be funny in terms of our projecting our subjective responses onto the world. But McDowell argues that they cannot give an account of these subjective responses without using the concept of amusement. This is problematic because concepts such as the concept of amusement pick out features of the world — they pick out things which have that property — and if the non-cognitivist wants to deny our judgements involve belief in such features then they need to be able to give an account of our judgements without using such concepts.

McDowell next goes on to claim that exactly the same is true in the case of moral sentiments — there is no way to identify which sentiments are moral ones without explaining them in terms of finding things to be good or bad. Blackburn attempts to explain the fact

that things seem good or bad in terms of our sentiments, but his position will be viciously circular if any account of moral sentiments requires a grasp of what it is for something to be good or bad.

(It's worth noting why this argument does not apply to *cognitivist* projectivist accounts, such as Mackie's error theory. Mackie is free to make use of moral concepts like *good* and *bad* when describing his position, it's simply that he holds that there are no things in the world which satisfy these concepts. Mackie holds that when we judge something to be good, say, we are not projecting some *attitude* onto the world but are simply projecting our belief that it is good onto the world. According to Mackie this belief is false, since nothing actually has the property of being good, but he agrees with the realist that the right account of moral concepts is the cognitivist one. Unlike the non-cognitivist, Mackie thinks that moral judgements involve a claim that a thing in the world has a moral property. And he has no need to characterise moral sentiments in terms which don't make use of moral concepts.)

Miller's response

Fundamentally McDowell's claim is that the only way we can identify moral sentiments is along the following lines:

r is a subjective response involved in S judging something to be good if and only if S has r any time she finds something to be good.

The right-hand side of this biconditional involves the very concept — that of things being good — which Blackburn's non-cognitivist projectivism seeks to explain. Blackburn attempts to explain the fact that things seem to be good in terms of our subjective responses to the world, but if McDowell is right and there is no way to identify moral subjective responses without using the concept of good then Blackburn's position is viciously circular. Blackburn seeks to explain the right-hand side of the above biconditional using its left-hand side, but, according to McDowell, the only available account of the left-hand side is the right-hand side, so Blackburn's explanation fails.

But why should we believe that this is the only way to characterise moral sentiments? Couldn't there be different ways of identifying whether a sentiment is a moral one without checking whether it involves finding something to be good? What the non-cognitivist needs to do is to fill in the right-hand side of the following biconditional without using the concepts of moral properties or moral sentiments:

r is a subjective response involved in S judging something to be good if and only if S has r any time she ...

Why is it McDowell thinks this cannot be done?

Miller's analysis of McDowell's argument focuses on McDowell's claim that "An inclination to laugh" is not a satisfactory answer: projecting an inclination to laugh would not *necessarily* yield an apparent instance of the comic" (McDowell, 1987/1998l, p. 158, emphasis added). Miller (2010) acknowledges that it may be true that there is no *necessary*, or analytic, definition of what it is for a subjective response to be moral, but he argues that that is not what Blackburn needs. He suggests we may be able to do the required work "on *a posteriori* explanatory grounds" (Miller, 2010, p. 333), and McDowell has given us no reason to believe that will not be possible:

This is clearly not a question that can be settled from the armchair. Can psychology provide us, in naturalistic terms, with some characterization of the sentiment of amusement? Or, in the ethical case, can psychology provide us, in naturalistic terms, with some characterization of distinctively ethical sentiments? McDowell's argument at best points to an explanatory space which the quasi-realist realizes that he has to fill. ... the crucial point is simply that McDowell has so far provided no argument, let alone a knock-down argument, against the thought that the quasi-realist can avail himself of a substantial naturalistic explanation of the nature of ethical sentiment. (Miller, 2013, p. 78).

And this seems reasonable; why not believe that psychology will not some day be able to give us a way to identify moral sentiments which doesn't rely on checking whether they involve finding something to be good? McDowell simply gives us no reason why this can't be done. Miller argues that all the non-cognitivist projectivist needs to be able to give an account of their position is to find a way to identify moral sentiments without using moral concepts, and this may well be possible after substantial empirical work.

A possible counter

Miller's response seems correct to the argument as McDowell presents it in "Values and Secondary Qualities" (McDowell, 1985/1998m, pp. 141–146) and "Projection and Truth in Ethics" (McDowell, 1987/1998l, pp. 157–162). McDowell asserts that the non-cognitivist can find no way to *identify* moral sentiments without using the concepts of moral properties, but he gives us no argument for this claim, and this claim seems to be limit on what *a posteriori* psychological research could accomplish. Why couldn't psychological research eventually find a way to distinguish between moral and non-moral sentiments? What justifies McDowell asserting, from the armchair, that it cannot?

However, I think there may be a better form of McDowell's argument, different to the one he presents here. Remember that what the non-cognitivist needs to be able to do is to explain our moral judgements without making use of the concepts of moral properties. They attempt to do this in terms of moral sentiments. Miller points out that McDowell gives us no reason to believe that there isn't a way to *identify* moral sentiments without

using concepts like good and bad. My concern here is that for the non-cognitivist to have an account of our moral judgements they need to do more than identify moral sentiments, they need to be able to *understand* them.

Blackburn's quasi-realist project is to earn our right to a notion of moral truth from the non-cognitivist projectivist starting point. The starting point provides an explanation of moral judgements, in terms of our moral sentiments. And then from his account of moral sentiments, and our attitudes towards them, he claims that we can earn our right to the notion of moral truth. (We saw this in §5.2 ("Quasi-realism") — he argues that we can earn our right to the notion of moral truth by evaluating attitudes towards moral sentiments and moral sensibilities. In brief, the idea is roughly that moral sentiments and sensibilities are true if our first-order moral theory should judge them to be appropriate.)

My suggestion here is that in order to do this work Blackburn needs to do more than simply identify when a sentiment is a moral one, he also needs to be able to understand moral sentiments — he needs to be able to grasp the sense of expressions referring to moral sentiments. Mere identification is not sufficient to do the kind of work his quasi-realist project sets out to do.

So, why might the non-cognitivist not be able to *understand* moral sentiments? Well, it turns out that this is a consequence of the first half of the no-priority view — the claim that *understanding the response requires understanding of the properties* — that we saw in the last chapter. McDowell thinks that this is true both for secondary qualities and also for values. He thinks that to understand an experience of something as seeming good requires some understanding of what it is for something to be good — grasp of the response that involves *finding something to be good* requires grasp of the concept of the property of goodness. Since moral sentiments involve the response of finding something to be good, this means that you can't understand moral sentiments without grasping the concepts of the properties of good and bad. And, as we saw above, this is precisely what the non-cognitivist account can't rely on, since they deny that our moral judgements involve belief in the properties of good and bad, and so they can't explain our sentiments in those terms — they need to be able understand our moral sentiment without making use of an understanding of moral properties.

When evaluating this premise in its use in McDowell's argument against the absolute conception in §4.2.2 I acknowledged that this half of the no-priority seems highly plausible, because it is very natural to characterise colour experience in terms of things seeming to be a particular colour. And so I concluded that perhaps McDowell is right to claim that we can't grasp the concept of the response without also grasping the concept of the property. If this is true then it can be used to save the contaminated response argument. This is because it implies that any account of moral sentiments requires one to understand the possibility of things having moral properties — to be a cognitivist about values. And this is precisely what Blackburn requires his austere starting point to avoid.

If this claim is true then McDowell's denial that we can *identify* moral sentiments without understanding the concepts of moral properties is simply unnecessary for him to undermine quasi-realism. All he needs to do is show that you can't fully *understand* moral sentiments (grasp the sense of expressions relating to moral experience) without already understanding moral properties. I noted in the previous chapter that the no-priority view is not a claim about limitations on what's needed to know the *extension* of the response, but on grasping its *sense*.

Of course, this version of the argument is still guilty of depending on a claim which McDowell has provided no argument for. In Miller's version of the argument McDowell simply assumed that:

There is no way of *identifying* moral sentiments without using the concepts of moral properties.

And in my suggested alternative version of the argument we need to assume that:

There is no way of *understanding* moral sentiments without using the concepts of moral properties.

However, the assumption required for this suggested amended version of the argument is significantly weaker than the one Miller identifies. That assumption involved a claim about the limitations of *a posteriori* psychological research, and seems deeply questionable. In contrast, the latter assumption is a claim about the nature of the concept of moral sentiments and experience (what is required to grasp the sense of expressions relating to moral sentiments and experience), and at the very least seems to be one which *a priori* philosophical work is appropriate for. However, although this claim seems plausibly true, McDowell has provided little in the way of argument for this claim and so it would be wrong to claim that this form of the contaminated response argument is a complete success.

5.4.3 Summary

Both the disentangling and contaminated response arguments are attempts to show that non-cognitivist projectivism can't give a satisfactory account of moral judgements without presuming an understanding of moral properties such as goodness, and thus conflicting with non-cognitivism (since if moral non-cognitivism is true then our understanding of moral talk is in terms of desires and sentiments, not the properties of items in the world). The disentangling argument does this by focusing on the features of the world which our moral judgements are responses to. It is the argument that there is no way to identify these items without relying on our prior understanding of the concept of the property of goodness. McDowell argues that the projectivist needs to be able to identify these features

because otherwise our moral judgements would not form a single conceptual practice — a case of “going on in the same way” rather than just reacting to the world at random. McDowell’s justifies the claim that the non-cognitivist can’t identify these features of the world using the uncodifiability thesis — the claim that “the best generalizations about how one should behave hold only for the most part” (McDowell, 1979/1998n, p. 58). Despite any rejection of this thesis striking McDowell as “merely fantastic” (McDowell, 1981/1998i, p. 215) the only passage approaching an argument to support this claim makes an unjustified leap from an epistemological claim to a metaphysical conclusion. So, the disentangling argument is not fatal for Blackburn’s non-cognitivist starting point. However, it does leave McDowell with an edge over Blackburn, since it demonstrates that for Blackburn’s starting point to be available the uncodifiability thesis must be false, whereas McDowell’s meta-ethics is compatible with the truth or falsity of the uncodifiability thesis. If we ended up endorsing a view which included the uncodifiability thesis then Blackburn’s meta-ethics would be dead in the water.

In the contaminated response argument McDowell turns his attention away from the the items in the world which our moral sentiments are responses to, and to the moral sentiments themselves. As it stands in the most straightforward reading of McDowell, this is the argument that we cannot find a way to identify moral sentiments — to distinguish moral from non-moral sentiments — without relying on our understanding of the concept of the property of goodness. However, as Miller rightly points out, McDowell is effectively making a claim about the limits of *a posteriori* science from the armchair. With next to no argument McDowell is effectively claiming that psychology could never identify when someone was having a moral experience. This is clearly a highly questionable claim.

In response I suggested that McDowell could run a very similar argument with a significantly less problematic premise — he doesn’t need to show you can’t *identify* moral sentiments without presuming an understanding of the moral properties, instead he only needs to justify the claim that you can’t *understand* moral sentiments without relying on your understanding of moral properties. This claim is the first part of the no-priority view that we saw in the last chapter — understanding the response requires understanding of the property — and when it came up there it seemed reasonable. However, McDowell has failed to provide a strong argument for this claim, and Blackburn is likely to simply reject it. So although in this version of the argument the key premise is much weaker than that required by the disentangling argument, it still fails to decisively demonstrate that non-cognitivist projectivism is an incoherent place for Blackburn’s meta-ethics to begin.

5.5 The second explanatory test: the explanatory failure of moral facts

We have seen that McDowell's arguments which attempt to show that Blackburn's quasi-realism fails the first explanatory test are inconclusive, and unlikely to convince Blackburn to drop his austere starting point. So, what about the second explanatory test, the test of whether value facts pull their weight in explanations of moral judgements? Blackburn argues that McDowell's moral realism fails this test, and accuses McDowell of merely citing moral truths without giving us any account of how they explain anything:

The objectivist illusion was (and I fear still is) to think that mentioning a moral reality, and flattering our understanding of it, affords some explanation of our practices in evaluation and judgement. The realist has no explanation of our evaluations, and he has no explanation of the structure of moral truth — such things as supervenience simply become brute mysteries. His illusion was that he had a theory, that by citing moral reality he could do away with the need to earn the concepts associated with objectivity. (Blackburn, 1981, p. 186).

Blackburn rejects McDowell's siblings metaphor (the no-priority view of values), arguing that there is “at best a one-way dependency, of property on sensibility” (Blackburn, 1993d, p. 377). He acknowledges that ethical properties may be dependent upon our ethical sensibilities, in that if our sensibilities were to change (if we had different value experience) then our value judgements would change. But he argues that there is no dependency in the other direction, “one would not see by any such means a naturalized dependency the other way [of sensibility on property]” (Blackburn, 1993d, p. 377). This leads him to conclude that “Moral ‘states of affairs’, above all, play no role in causing or explaining our attitudes, their convergence, their importance to us.” (Blackburn, 1981, pp. 185–186), and so he endorses projectivism and claims that we should understand our ethical commitments as “[filling] other cognitive functions than describing the way of the world” (Blackburn, 1993d, p. 378).

As we saw when looking at McDowell's response to Mackie's argument from epistemological queerness in §4.3.2 of last chapter, McDowell agrees that a straightforward intuitionistic realism is guilty of expanding reality by mere postulation, and of simply introducing cognitive faculties using vague analogies with sensory perception, but with no account of how they enable us to know facts about values:

The contrast — the unearned employment of the notion of truth that projectivism rejects — is a position that expands reality by mere postulation, beyond what the projectivist is comfortable with, to include an extra population of distinctively value-involving states of affairs or facts. Corresponding to this, it purports to equip us with special cognitive faculties by whose exercise we become aware of this special field of knowable fact. These special cognitive

faculties are vaguely assimilated to the senses, but no detailed account can be given of how they operate, such as might make it clear to us — as clear as it is in the case of the senses — how their exercise affords us access to the relevant range of circumstances. The assimilation to the senses gives this intuitionistic position the superficial appearance of offering an epistemology of our access to evaluative truth, but there is no substance behind this appearance. (McDowell, 1987/1998l, pp. 153–154).

And he states that, if we understand realism as simple intuitionistic realism, then Blackburn would be right:

Realism shirks the obligation [to earn truth in ethics], and the clear alternative is projectivism. (McDowell, 1987/1998l, p. 155).

However, McDowell argues that Blackburn is wrong to assume that once we reject the blunt intuitionistic realism the only remaining option is projectivism. Instead, he urges that we should draw an analogy between values and secondary qualities, adopt the siblings view, and recognise that values are “unintelligible except as modifications of a sensibility like ours” (McDowell, 1985/1998m, p. 143). And to endorse the secondary-quality account of values is “to give up the idea that [value]¹⁴ itself, were it real, would need to be intelligible from a standpoint independent of the propensity to [feel moral responsibility]” (McDowell, 1985/1998m, p. 146).

If McDowell’s no-priority view is correct, and value is unintelligible from the standpoint independent of the propensity to have moral responses to items in the world, then that would explain why value facts may seem, to Blackburn, to pull no weight. McDowell’s no-priority view means that both value facts and our responses to values are only intelligible for someone who’s had the appropriate upbringing. From inside that point of view we can make sense of the fact that we are able to perceive moral facts, but they will not be intelligible from a point of view which doesn’t presume our understanding of values. Blackburn starting point attempts to explain the role of value facts, but without presuming any of our preexisting understanding of moral responses and value facts. And so (if McDowell is correct) it should come as no surprise that from Blackburn’s starting point we cannot see how value facts pull any explanatory weight.

5.5.1 Blackburn’s challenge

Even if this is correct, it isn’t particularly satisfying. McDowell has given us a reason why Blackburn’s starting point fails to recognise the role value facts play in explaining our moral judgements, but he has not given us any reason to believe that value facts do

¹⁴In the actual passage McDowell is using an analogy with fear to explain his account of values. The original version is as follows: “to give up the idea that fearfulness itself, were it real, would need to be intelligible from a standpoint independent of the propensity to fear” (McDowell, 1985/1998m, p. 146).

play such a role. Blackburn argues that McDowell has failed to give us any justification for talking in terms of us perceiving moral facts, rather than their being projected onto the world. The problem with McDowell's view is that "there is no theory connecting these truths to devices whereby we know about them" (Blackburn, 1985/1993b, p. 163). According to Blackburn:

We speak of perception whenever we think of ourselves as properly indicating the truth: in other words, whenever we feel able to say that 'if it hadn't been the case that p , I would not be committed to p '. But this is not the end of epistemology, but its beginning, for the theorist's job is to reflect upon our rights to hold such conditionals. ... But genuine cases of perception standardly demand stories with different ingredients. 'If it hadn't been the case that the shape was square, I would not have believed that it was' can be said because we are causally affected by shapes and can use those effects to deliver verdicts on them. (Blackburn, 1985/1993b, p. 161).

Blackburn's challenge is to explain our right to hold the following conditional, which I'll refer to as the Key Conditional:

The Key Conditional: "if it hadn't been the case that p , I would not be committed to p " (Blackburn, 1985/1993b, p. 161).

He argues that to be justified in talking in terms of *perceiving* moral facts we need to be able to tell the kind of story we can in other cases of perception, such as the causal story we can tell about our visual perception of shapes. In the case of our experience of a shape we can explain how we are able to know that an object has a particular shape — because light reflected off the surface of the object is detected by our eyes, and we are then able to recognise that the object is square, say. And if the object's shape was different then a different pattern of light would have been reflected, leading us to make a different judgement about its shape. This explains our right to hold the Key Conditional in the case of primary quality experience. Blackburn argues that we can tell no such story in the case of the perception of values, so we can't justify the claim that value facts play any role in causing our beliefs. This means that we should be projectivists, rather than realists, about values.

5.6 The epistemology of susceptibility to reasons

But as we saw in the previous chapter, McDowell doesn't endorse a simple perceptual picture of our ability to know moral facts. In this respect he thinks that the analogy between values and secondary qualities should be rejected:

The disanalogy [between values and secondary qualities], now, is that a virtue (say) is conceived to be not merely such as to elicit the appropriate "attitude"

(as a colour is merely such as to cause the appropriate experiences), but rather such as to *merit* it. And this makes it doubtful whether merely causal explanations of value experience are relevant to the explanatory test . . . we should be raising that question about explanations of a different kind. (McDowell, 1985/1998m, p. 143, emphasis in original).

And in “Projection and truth in ethics” he adds further clarification:

Instead of a vague attempt to borrow the epistemological credentials of the idea of perception, the position I am describing aims, quite differently, at an epistemology that centres on the notion of susceptibility to reasons. (McDowell, 1987/1998l, p. 162).

So, how does this talk of “susceptibility to reasons” (McDowell, 1987/1998l, p. 162) enable McDowell to explain his justification for endorsing the Key Conditional? His answer is that all he needs to do is “[establish] that one would not have arrived at the belief that *p* had it not been for good reasons for it” (McDowell, 1987/1998l, p. 163), and there is nothing wrong with “the excellence of the reasons [being] vindicated from within the relevant [moral] way of thinking” (McDowell, 1987/1998l, p. 163).

McDowell claims that all that is needed to respond to Blackburn’s challenge is to show that if we didn’t have good *reasons* for believing that *p* then we would have not arrived at that belief. There is no need to be able to tell an epistemological story along the same lines as we would for the case of our perception of shapes, or even colours; all we need to do is show that our belief is dependent upon our having good reasons for it — our moral attitudes must be merited.

Blackburn’s prejudice about justification

McDowell suggests that Blackburn doesn’t see the demonstration of the existence of *reasons* alone as sufficient for responding to his challenge because of a prejudice about the nature of justification:

[Blackburn’s “metaphysical understanding” fixes] a conception of the kinds of cognitive occurrence that can constitute access to facts; nothing will serve except what can be conceived in terms of the impact on us of the world as the “metaphysical understanding” understands it. That is why one is not allowed to count as protecting one’s right to a conditional of the form “If it had not been that case that *p*, I would not have become committed to the belief that *p*”, if one establishes that one would not have arrived at the belief that *p* had it not been for good reasons for it, with the excellence of the reasons vindicated from within the relevant way of thinking. (McDowell, 1987/1998l, p. 163).

Blackburn’s mistake is to presume that our right to hold the Key Conditional — “If it hadn’t been the case that *p*, I would not be committed to *p*” (Blackburn, 1985/1993b, p. 161) — must be supported by developing an epistemological story along the lines of

that of primary quality perception. McDowell argues that we shouldn't presume that this right can always be justified in the same way, and that although for values we can't tell a story along those lines — such as a story about light being reflected by objects in the world, our eyes detecting that light, and the pattern that reaches our eyes being determined by the shape of the object — that doesn't mean that we don't have the right to talk about our beliefs being determined by the facts, or being justified.

5.6.1 Earning truth in ethics

So now the question remains, if we can't tell the kind of epistemological story we can in the case of primary quality perception, how do we earn our right to talk of detecting truths in ethics? According to McDowell the answer is that we can earn our right to talk of the detection of moral truths simply by providing arguments along the lines of the arguments for imminent realism we saw in §5.3.1. There we saw Blackburn identify two features of truth that imminent realists, like McDowell, think we need to demonstrate:

It should be unique, and it should stand at sufficient distance from our actual opinions for us to entertain the thought that our actual opinions may not be true. (Blackburn, 1993d, p. 370).

I'll next look at both of these, and explain McDowell's arguments that these are satisfied in the case of ethics. Finally I'll draw a parallel with mathematics, making use of the conclusions I reached towards the end of the first chapter, to give us some reason to believe that McDowell has genuinely earned our right to the Key Conditional.

The distance between truth and our opinions

I'll start with the second of these features, that the truth and our opinions can come apart. McDowell gives two reasons to believe that this is the case in ethics, the first of which is that we can make sense of the idea of truths that we are not aware of. Consider the case of colours, clearly there can be truths of which we are not aware: "An object's being such as to look red is independent of it actually looking red to anyone on any particular occasion; so . . . an experience of something as red can count as a case of being presented with a property that is there anyway — there independent of the experience itself" (McDowell, 1985/1998m, p. 134)). And he makes the same claim about values: "Values are not brutally there — not there independently of our sensibility any more than colours are: though, as with colours, this does not prevent us from supposing that they are there independently of any particular apparent experience of them" (McDowell, 1985/1998m, p. 146). This seems correct, we can make perfect sense of the idea that we have a moral obligation to act a certain way, even if no one has happened to notice it.

The other reason for believing that truth and our opinions can come apart that McDowell gives is the fact that it's possible for us to be mistaken in our moral judgements. To

justify this claim he points out that moral judgements can be highly contentious and that people change their minds about which moral judgement they should endorse. This implies that, if there are moral facts, we may sometimes go wrong in our moral judgements, and so we are required to be able to rationally evaluate our own value judgements:

awareness that values are contentious tells against an unreflective contentment with the current state of one's critical outlook, and in favour of a readiness to suppose that there may be something to be learned from people with whom one's first opinion is to disagree. . . . a sensible person will never be confident that his evaluative outlook is incapable of improvement (McDowell, 1985/1998m, p. 145).

The knowledge that our moral judgements can be wrong, which is brought out by our awareness of the controversy of many moral claims, shows that we have an understanding of moral truth as independent from what we currently believe. This does not necessarily demonstrate that our current moral beliefs are false, but at least shows that we acknowledge that possibility.

The uniqueness of the truth

However, the contentiousness of value judgements may appear to undermine the other feature of truth that Blackburn identifies as required by the imminent realist, that of the uniqueness of truth. As we saw in the previous chapter, Mackie's argument from relativity attempts to use precisely this to argue against moral realism. McDowell's reply (which we saw in §4.3.1) was that contentiousness alone doesn't show that there are no moral facts. What we need to do in the face of disagreement is to provide rational reasons to justify endorsing one particular moral view over another:

Earning the notion of truth . . . is a matter of arguing that we do after all have at our disposal a conception of reasons for ethical thinking that is sufficiently rich and substantial to mark off rationally induced improvements in ethical stances from alterations induced by merely manipulative persuasion. (McDowell, 1987/1998l, p. 156).

If we can do this then we can justify believing that truth is unique. Even though we may disagree about what the truth is, as long as we can give rational reasons for preferring one view over the others then that does not undermine the claim that the truth is the view that we have best reasons for.

5.6.2 The parallel with mathematics

So, McDowell concludes that we can earn our right to the notion of truth in ethics. Our right to Blackburn's Key Conditional, "if it hadn't been the case that p , I would not

be committed to p " (Blackburn, 1985/1993b, p. 161), is provided by the fact that if we didn't have good reasons for p then we would not be committed to the belief that p .

But there may still seem to be something spooky about an epistemology that is supported by our susceptibility to reasons, rather than by a story parallel to that which we can provide for primary quality perception. To attempt to dispel this sense of spookiness I shall compare the case of ethics with that of mathematics, where our knowledge of the facts is also supported by reasons alone, and no perceptual story can be told.¹⁵

Consider Pythagoras' Theorem. When asked what justifies our right to hold the Key Conditional we would say that if we didn't have a successful proof for Pythagoras' Theorem then we would not be committed to it. The proof provides our *reason* for believing in the theorem, and that plays precisely the same role as our moral reasons do in the case of ethical claims.

In the case of mathematics we know that there are mathematical facts because we are able to prove theorems. In the case of ethics we may not be able to provide infallible proofs, but, as we saw in §3.5 of chapter 3 ("Knowledge of meaning and epistemic fallibilism"), McDowell argues that the fact that an epistemic capacity is fallible does not prevent it from enabling us to know truths when all goes well. According to McDowell's epistemic fallibilism it is possible to know facts in a particular domain even if your ability to know such facts is fallible and sometimes goes wrong. We would expect our ability to know moral facts to be fallible in this way, in that sometimes we will make mistakes in our moral judgements, but that doesn't undermine our claim to sometimes make correct moral judgements.

In the crucial sense there is no significant difference between mathematics and ethics here. In both cases we cannot tell any substantial perceptual story for how we are able to know the truths, but instead all we can refer to is the fact that we have good *reasons* for our beliefs. In mathematics this doesn't prevent us from claiming that we can know mathematical truths, so why shouldn't we say the same in the case of ethics?

The justification for realism in mathematics

Another interesting parallel between ethics and mathematics is between the arguments we saw for the claim that we could earn truth in ethics, and the argument for moral realism I presented at the end of the first chapter.

To earn truth in ethics we needed to show that it is unique, and that the truth and our opinions can come apart. In the case of mathematics I argued that we understand

¹⁵In his critical notice of Williams' book *Ethics and the Limits of Philosophy* McDowell himself briefly notes the parallel between ethics and mathematics, saying that in both discourses "The fact impresses itself by way of the reasons alluded to: one might ask how else it could do so, if not by way of reasons." (McDowell, 1986, p. 379). In this section I am effectively fleshing out that parallel and building on it by drawing on the argument for realism about mathematics I presented at the end of the first chapter.

truth to be potentially evidence-transcendent, since there is mathematical reasoning which provides a fallible access to truths. Although it would be wrong to say that we understand moral truths to be potentially evidence-transcendent (since it would conflict with the thesis that “ought implies can” if we could have moral obligations which we can’t know about), it’s useful as another example of an internal argument for realism.

At the end of chapter 1 I argued that McDowell was wrong to claim that our understanding of truth in mathematics is evidence constrained and to think that “There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations.” (McDowell, 1989/1998g, p. 354). To refute this I argued that our best understanding of the epistemology of mathematics essentially involves the idea that mathematical assertions can have potentially *evidence-transcendent* truth-conditions. I justified this claim by pointing out that when assessing the plausibility of conjectures and planning proofs mathematicians engage in reasoning which is fallible; their mathematical training enables them to use intuition and speculation to provide good reasons for thinking that a conjecture is likely to be true. These reasons are unlike proofs in that they are fallible and sometimes go wrong, which shows that we can understand how mathematical truths can come apart from our beliefs, and that in mathematics, *contra* McDowell, “we can contemplate in thought a reality that is determinate *beyond our access to it*” (McDowell, 1989/1998g, p. 355, emphasis in original).

Similarly, McDowell’s argument that we can earn our right to the notion of truth in ethics involved the demonstration that we can make sense of the distinction between the truth and our opinions. He did this by pointing out that we believe that our opinions will sometimes be wrong. We are aware of the contentiousness of some of our moral judgements, and this is valuable because it reminds us of the possibility our judgement could be wrong:

awareness that values are contentious tells against an unreflective contentment with the current state of one’s critical outlook, and in favour of a readiness to suppose that there may be something to be learned from people with whom one’s first opinion is to disagree. . . . a sensible person will never be confident that his evaluative outlook is incapable of improvement (McDowell, 1985/1998m, p. 145).

And so he concludes that we can make sense of moral facts being independent of what we happen to think.

Both these arguments are cases of internal arguments for endorsing a particular understanding of truth in that domain, and if there is nothing spooky about my argument that truth in mathematics is potentially evidence-transcendent, then neither should there be anything spooky with McDowell’s argument that we can earn our right to truth in ethics.

5.7 Conclusion

In the last chapter we saw that if we endorse McDowell's no-priority view of values and draw on the analogy between values and secondary qualities then we can give an account of the reality of values which is much less queer than Mackie makes them seem. In this chapter McDowell attempted something more ambitious, not merely to present an account of values which isn't queer, but to undermine Blackburn's account by arguing that Blackburn's starting point is incoherent because it is inconsistent with denying the reality of values. This was the first of the two explanatory tests which I highlighted as being at the heart of the debate between McDowell and Blackburn, the test of whether the explainer can consistently deny the reality of values. McDowell attempted to demonstrate that Blackburn failed this test but I concluded that both his arguments, the disentangling and contaminated response arguments, were inconclusive.

The disentangling argument failed to demonstrate the incoherence of Blackburn's position because it relied on the claim that morality is uncodifiable — the claim that there is no way to explain how one should live in non-moral terms. McDowell fails to provide any significant evidence for this claim, and so Blackburn can simply deny it. However, the disentangling argument does leave McDowell's metaethical position with a slight edge over Blackburn's, since it's still up for debate whether or not the uncodifiability thesis is true. If it turns out to be true then Blackburn's starting point would be unavailable, because he will be unable to explain how moral judgements form part of a genuine conceptual process as there is nothing in the world which correct moral judgements have in common other than being a response to the property of goodness.

The success of McDowell's contaminated response argument is equally limited. The argument as McDowell presents it seems to rely on a premise which is completely unreasonable. As Miller points out, McDowell seems to simply assert that there is no way to identify moral sentiments without using the concepts of moral properties, but this is a claim about the limitations of *a posteriori* empirical research made from the armchair and with no justification; why couldn't psychology find a way to identify moral sentiments? However, I suggested that we can modify McDowell's argument to be the claim that there is no way to *understand* moral sentiments without the concepts of moral properties. This is the first half of the no-priority view that we saw in the last chapter, the claim that understanding the response requires understanding of the properties. This seems to be a much more reasonable claim, and one which could potentially be settled from the armchair, but McDowell provides little in the way of justification for it and so it seems that, for now, Blackburn can simply reject it.

Blackburn argued that we should reject McDowell's position because it fails the other explanatory test, the test of whether value facts pull their weight in the explanation. Blackburn argues that McDowell can't justify his right to endorse the conditional, which

I call the Key Conditional, “if it hadn’t been the case that p , I would not be committed to p ” (Blackburn, 1985/1993b, p. 161). However, McDowell replies by arguing that Blackburn only thinks that he can’t earn his right to this conditional because of a prejudice about the nature of justification. McDowell argues that he can explain his right to this conditional by claiming that if p were false then we wouldn’t have good *reasons* for p . I draw a parallel with mathematics to demonstrate how reasonable an epistemology of susceptibility to reasons in. In the case of mathematics cannot tell a substantial perceptual story for how we are able to know the truths; instead all we can refer to is the fact that we have good *reasons* for our beliefs. I pointed out that in mathematics this doesn’t stop us from claiming that we are able to know mathematical truths, or claim our right to hold the Key Conditional, so why shouldn’t we say the same in the case of ethics? I also drew an analogy between McDowell’s arguments for realism in ethics (the demonstration that the truth is unique and that truth and our opinions can come apart) and my arguments for mathematical realism at the end of the first chapter (the demonstration that mathematicians sometimes use fallible epistemic practices to work towards knowing the truth). Both of these are cases of internal arguments for endorsing a particular understanding of truth in that domain, and if there is nothing spooky about my argument for endorsing a potentially evidence-transcendent understanding of truth in mathematics, then neither is there anything spooky with McDowell’s argument that we can earn our right to truth in ethics.

In summary, Blackburn and McDowell have both failed to fatally damage each other’s positions. And so my conclusion must be inconclusive. McDowell holds the upper hand in that Blackburn’s position is a hostage to fortune in two respects: if we conclude that morality is uncodifiable or that we cannot understand moral responses without understanding moral properties, then his starting point would be unavailable. But on the other hand if we concluded that McDowell’s no-priority view is a bad understanding of our ethical concepts then he would be unable to save McDowell’s form of moral realism from Mackie’s arguments from queerness.

Conclusion

In the first chapter I concluded that McDowell's response to the manifestation argument is a success, but only if we endorse his claim that our theory of meaning must be modest, and his epistemological fallibilism. Modesty involves the claim that the meaning of our sentences can itself be directly audible when we speak: "[the content of sentences as not] something "beneath" the words, to which we are to penetrate by stripping off the linguistic clothing; rather, as something present in the words — something capable of being heard or seen in the words by those who understand the language" (McDowell, 1987/1998e, p. 99). If this is correct then McDowell is right to claim that we can manifest our knowledge of what a sentence can be used to assert simply by asserting the sentence itself, an essential premise in his response to the manifestation argument.

But merely being able to manifest our understanding of sentences by speaking them wasn't enough to justify the realist conclusion; I also needed to deal with a potential ambiguity in the second premise of McDowell's argument. I needed to demonstrate that we understand our sentences to have potentially evidence-transcendent truth-conditions, and not evidence-constrained truth-conditions. For this I brought in McDowell's arguments for fallibilism.

Fallibilism is the claim that we are able to gain knowledge of x even in cases which we may not be able to distinguish from cases in which x is false. For example, it is the claim that it is possible to know, by looking, that there is a tomato in front of you even if you haven't completely ruled out the possibility that the object in front of you is actually a mere façade. I argued that if a fallibilist epistemology is appropriate for a particular discourse then that means that in that discourse "we can contemplate in thought a reality that is determinate *beyond our access to it*" (McDowell, 1989/1998g, p. 355, emphasis in original). And so we must understand truth in that domain to be potentially evidence-transcendent. This dealt with the ambiguity in McDowell's second premise which Wright's arguments demonstrated.

Although much of the work on fallibilism which I draw on came from the paper "Mathematical Platonism and Dummettian Anti-Realism" (McDowell, 1989/1998g) in which McDowell endorses anti-realism about mathematics, I argued that McDowell was wrong to do so because he is wrong to claim that "There is nothing for linguistic competence to show itself in, in mathematics, apart from responses to proofs and refutations."

(McDowell, 1989/1998g, p. 354). I argued that when mathematicians assess the plausibility of conjectures, and when planning the developing of mathematical proofs, their reasoning fits a fallibilist epistemological picture. In the case of conjectures for which we currently lack a proof or refutation mathematicians report having evidence for a conjecture's truth which they feel justifies their belief in it, and yet falls far short of the absolute certainty provided by a proof. And when planning a mathematical proof mathematicians rely heavily on intuition and speculation, an epistemic methodology which may be a guide to truth but is always capable of going wrong. So I concluded that, *contra* McDowell, we should be realists about mathematics.

McDowell's response to the rule-following problem relied on the rejection of the master thesis, instead he claimed that things can come to mind which are capable of "[imposing] a sorting of extra-mental items into those that accord with it and those that do not" (McDowell, 1993/1998h, p. 270) — things can come to mind that can have normative significance without requiring interpretation. He didn't give us an account of *how it is possible* for meanings to come to mind, but instead demonstrated that this view is simple common sense, and that if we believed that anything that came to mind required interpretation to have meaning then we would be led into an infinite regress which makes meaning impossible.

To make McDowell's response to Kripke's problem look less like he was simply insisting that the problematic thing was possible and then leaving it at that, I looked at how it fits into his overall view of the world and our experience of it. McDowell argues that the world, and our experience of it, must be conceptually structured (like our mental life). I showed how his arguments for this conclusion relied on his endorsement of internalism about justification and epistemic fallibilism, neither of which I evaluated. But if these are true then McDowell's view of reality seems necessary, if knowledge is to be possible at all, and his view of the inner fits well with that.

However, when looking at McDowell's response to the seems right / is right distinction I revealed a problem with his account of the inner: McDowell claims that inner objects "have no existence independently of [the awareness of "inner experience"]" (McDowell, 1994/1996, p. 21), but this makes it impossible for him to make sense of the possibility of our making certain mistakes about inner items, such as having intentions of which we are not aware. I argued that he should reject this account of the inner, but this leaves him with the need to draw the distinction between the inner and outer worlds.

In the first of the chapters on metaethics I concluded that McDowell had successfully presented an account of moral realism which isn't vulnerable to Mackie's queerness arguments. I explained the analogy McDowell draws between values and secondary qualities by drawing on the no-priority view, which comes up in a paper where he's engaged with Blackburn, not Mackie. This is the view that our understanding of moral properties and responses is co-dependent: understanding moral responses (sentiments and experience) requires understanding of the properties, and understanding the properties requires understanding of the response. This enables him to respond to the argument from relativity by pointing to the role our way of life plays in our moral responses, and to undermine the argument from metaphysical queerness, because if colour can only be understood in terms of how it affects certain sentient beings then there is nothing queer about the fact that values can only be understood in terms of how they affect us (in motivating us to act). And in response to Mackie's arguments from epistemological queerness I explained McDowell's Neurathian picture of our access to truth, why we shouldn't expect to be able to understand the justification for moral reasoning from outside, but how we are able to justify our right to talk in terms of moral perception and the detection of moral properties once we have had sufficient training. This response required the claim that when there is moral disagreement we can justify endorsing the position we do by giving rational reasons for preferring it, and if we find we cannot do this then we should reject moral realism since we cannot claim truth to be unique.

In the final chapter I argued that neither of McDowell's arguments which attempt to demonstrate that Blackburn's starting point is incoherent succeed. Both his contaminated response and disentangling arguments rely on premises which Blackburn can simply deny. But if we were to conclude that one of these premises was true — that either morality is uncodifiable and there is nothing which moral items share other than the fact that they are good or bad; or that we cannot understand moral sentiments other than in terms of their being responses to moral properties — then Blackburn's starting point would be unable to give its explanation of our moral sentiments without presuming cognitivism about values, and so would be unavailable. In this respect Blackburn's position is a hostage to fortune, but as it currently stands we have scant evidence for the truth of either of the required premises, and so we shouldn't give up on the quasi-realist project just yet. However, Blackburn's attacks on McDowell also fail because he presumes that our justification for beliefs must always fit a model which is justified by our analysis of our perception of the facts, and not simply by our being able to demonstrate that we have sufficient *reasons* for our beliefs. Finally, I developed a parallel between values and mathematics, drawing on what I said about mathematics at the end of the first chapter, to dispel any appearance of spookiness which an epistemology in terms of susceptibility to reasons may appear to

have.

Is there a unified notion of realism?

One could question whether there is a unified notion of realism which all of these conclusions share. To answer this question I'll swiftly go through the conclusions we reached on of the topics of this thesis, and then turn to look at what they have in common.

In chapter 1 we saw McDowell respond to the manifestation argument and argue that our sentences can have potentially evidence-transcendent truth-conditions, i.e. our sentences are capable of being true or false even in cases in which we don't have evidence for their truth-value (or a finite method guaranteed to detect their truth-value). A fundamental step in the argument for this conclusion was the claim that "Knowledge of what a sentence can be used to assert is knowledge which can be directly manifested . . . by using the sentence . . . to assert precisely that." (McDowell, 1981/1998b, p. 321). In chapters 2 and 3 we saw McDowell use a similar claim to respond to the rule-following challenge: according to McDowell our access to the facts about a sentence's meaning are accessible to us simply because we are able to "hear someone else's meaning in his words" (McDowell, 1984/1998o, p. 253). Unlike Wright, McDowell does not attempt to give an account of what constitutes the facts about meaning, and so when asked "What fact makes it the case that Jones means plus by '+'?" McDowell's answer can be simply "The fact that Jones means plus by '+'."

The idea that we are able to "hear someone else's meaning in his words" (McDowell, 1984/1998o, p. 253) cropped up again in chapter 4, where we saw McDowell drawing an analogy between learning a language and learning to be able to perceive the moral facts. Just as learning a language enables us to hear something new when we hear spoken language, so a decent moral upbringing enables us to be able to perceive the moral properties of actions.

In chapter 5 we saw McDowell's moral realism come under attack from Blackburn. Blackburn argued that McDowell's endorsement of realism must be wrong because his account failed to satisfy what I called the "Key Conditional": we should reject McDowell's moral realism because it can't explain our right to claim that "if it hadn't been the case that p ,¹⁶ I would not be committed to p " (Blackburn, 1985/1993b, p. 161). Blackburn argues that McDowell needs to do more in order to justify talking about moral facts being perceived rather than being merely projected onto the word. McDowell responds by arguing that Blackburn is presuming that we need to be able to *explain* our ability to know moral facts — explain how it is possible for us to perceive moral facts — but that this presumption is wrong. Instead all that is needed to defend realism is to justify

¹⁶Where p is a moral fact.

holding that the moral facts are independent of what we happen to think, and therefore are not merely our own projections.

What's common in all these cases is that McDowell is asserting that there are certain facts — potentially evidence-transcendent facts; facts about meaning; and moral facts — even though we can't give a detailed account of what constitutes such facts, or an explanation how we are able to know them. This quietist aspect of McDowell's philosophical approach was most apparent in the chapters on meaning and rule-following. As we saw above, when asked "What fact makes it the case that Jones means *plus* by '+'?" McDowell will simply reply "The fact that Jones means *plus* by '+'." And when asked how are we able to know what someone's spoken language means, the answer be that we can "hear someone else's meaning in his words" (McDowell, 1984/1998o, p. 253). Central to quietism is the rejection of certain "how possible" questions, and the attempt to instead undermine the assumptions which make it seem as though an answer to such questions is needed:

The right response to "How is meaning possible" . . . is to uncover the way of thinking that makes it seem difficult to accommodate meaning and intentionality in our picture of how things are, and to lay bare how uncompulsory it is to think in that way. (McDowell, 1993/1998h, p. 272).

In the case of meaning, McDowell did this by demonstrating that the master thesis is false. He argued that things in mind don't need to be interpreted to have normative significance, and once we accept this conclusion there is no need to worry about how it is possible for meanings to come to mind. After demonstrating that the seeming need for an explanation of our ability to know certain facts was based on a mistake McDowell concludes that we can rest easy without such an explanation.

Similarly, when responding to Blackburn's argument against moral realism McDowell doesn't present us with an *explanation* of our ability to know moral facts. In the case of our ability to know moral facts we can't tell a story along the lines of that which could be told for visual perception (such as a story about light being reflected by objects in the world, our eyes detecting that light, and the pattern that reaches our eyes being determined by the shape of the objects). But that doesn't mean that we can't justify the claim that we can have good reasons for our moral beliefs. Instead of providing an explanation for our ability to know moral facts, McDowell argued that all we needed to do was show that there is a distinction between truth and our opinions, and that the moral truths are unique — whenever there is a case of moral disagreement we can give reasons for endorsing one moral belief over another.

This explains why McDowell's style may sometimes seem oblique. McDowell doesn't tend to provide explanations of particular domains of truth, or explanations of our ability

to know certain things. Instead he aims to undercut the arguments which make it seem that such explanations are required. Once he has freed us from the picture which makes it seem that such an explanation is needed he concludes that his job is done. As we saw in the rule-following arguments, unlike Wright, McDowell doesn't take sides in arguments about how meaning is possible, instead he merely aims to show that the argument is based upon a misunderstanding. Once this misunderstanding is cleared up it will become clear that there is no need for an account which explains the possibility of meaning.

This means that it would be wrong to say that there is a unified notion of realism running through the three topics I've covered in this thesis. Or at least, there is no unified *explanation* of what makes certain facts true, or what makes it possible for us to know about them, because McDowell rejects the entire project of giving such explanations. Instead, the unified theme running through these topics is one of philosophical approach: to seek out the mistakes which make such explanations seem to be required, and to reveal that no such explanation is needed for us to believe in our ability to know about such facts.

Suggestions for further work

Another aspect of McDowell's philosophical approach which time and again appeared to be key in his arguments was the central role epistemological claims played. His response to the manifestation argument required that we endorse epistemic fallibilism, both in order to demonstrate that we have a realist understanding of truth, and to make a modest theory of meaning tenable.

McDowell's response to the rule-following challenge essentially involved two claims: the insistence that we are able to "hear someone else's meaning in his words" (McDowell, 1984/1998o, p. 253); and the claim that we are able to have conceptually structured experience of our inner states in order for them to have normative force without interpretation. And again, the feasibility of these claims required us to be fallibilists about justification, because we cannot rule out the possibility that we have made a mistake when listening to someone speak, or when remembering what we meant. He argued for these two claims by showing that if we deny them then we make meaning impossible, and I supported his position by drawing analogies with his general view of the world and our experience of it. This overall view of the world was justified by epistemological arguments, which essentially claimed that if we are epistemic internalists (if we need to have rational access to the warrant for our perceptual beliefs for our beliefs to be justified) then we *must* view the world and our experience of it as conceptually structured, otherwise knowledge would be impossible.

McDowell's account of moral realism also makes significant epistemological claims. He argues that by having the appropriate upbringing we gain the ability to perceive the moral

facts, but doing so requires us to be able to justify our moral view of the world. I drew an analogy between how upbringing enables us to see moral facts and how being taught a language enables us to hear what someone means from the sound of their words. Both are perceptual capacities which are learnt, both are fallible, and both seem to involve our perceiving something which supervenes on other properties (in the case of language by hearing sounds we are able to perceive meaning, and in the case of ethics by seeing acts we are able to see their moral properties). This is a very interesting aspect of McDowell's position, particularly in the relationship between his talk of moral perception and the role we must be able to give reasons for having the ethical sensibility which we do, and much more work would be needed to flesh this out and decide whether it is a good way of thinking about our access to moral facts.

Another recurring epistemological theme was that of McDowell's endorsement of direct realism. This cropped up in the first chapter where we saw McDowell argue that we have a fallible ability to directly know facts such as the fact *that there is a tomato* simply by looking. According to this view, "the fact itself is directly presented to view" (McDowell, 1982/1998d, p. 387). Our experience itself is an experience *that there is a tomato* — the fact *that there is a tomato* is an aspect of our experience. This claim played an essential role in McDowell's epistemic picture, and enabled him to argue that we have a fallible ability to know about the world, and so justify the claim that we understand sentences to have potentially evidence-transcendent truth-conditions.

Direct realism appeared again in McDowell's account of meaning. The rejection of the view that someone's speech needed interpretation in order for us to know what they mean, and the endorsement of the view that we can *directly* hear someone's meaning in their words is itself another case of McDowell's endorsement for direct realism.

It would be valuable to look in-depth at the debates surrounding direct realism, fallibilism and epistemic internalism. As we've just seen, all are essential for McDowell's position, and play key roles in justifying his arguments. In this thesis I have only attempted to give a brief sketch of McDowell's justification for endorsing these views, and taking a detailed look at the literature on these claims could well generate exciting insights both on McDowell's view and, if he is right, on the correct view of reality in general.

References

- Aristotle. (1925). *The nicomachean ethics* (D. Ross, Trans.). Oxford: Oxford University Press.
- Ayer, A. J. (1936). *Language, truth, and logic*. London: Gollancz.
- Blackburn, S. (1981). Reply: rule-following and moral realism. In S. Holtzman & C. Leich (Eds.), *Wittgenstein: to follow a rule* (pp. 163–187). London: Routledge & Kegan Paul Ltd.
- Blackburn, S. (1984a). The individual strikes back. *Synthese*, 58, 281–302.
- Blackburn, S. (1984b). *Spreading the word: groundings in the philosophy of language*. Oxford: Oxford University Press.
- Blackburn, S. (1988). How to be an ethical antirealist. *Midwest Studies in Philosophy*, 12(1), 361–375.
- Blackburn, S. (1991). Just causes. *Philosophical Studies*, 61(1), 3–16.
- Blackburn, S. (1993a). Attitudes and contents. In *Essays in quasi-realism* (pp. 182–197). Oxford: Oxford University Press. (Original work published 1988)
- Blackburn, S. (1993b). Errors and the phenomenology of value. In *Essays in quasi-realism* (pp. 149–165). Oxford: Oxford University Press. (Original work published 1985)
- Blackburn, S. (1993c). Moral realism. In *Essays in quasi-realism* (pp. 111–129). Oxford: Oxford University Press. (Original work published 1973)
- Blackburn, S. (1993d). Realism, quasi, or queasy? In J. Haldane & C. Wright (Eds.), *Reality, representation, and projection* (pp. 365–383). Oxford: Oxford University Press.
- Blackburn, S. (1993e). Supervenience revisited. In *Essays in quasi-realism* (pp. 130–148). Oxford: Oxford University Press. (Original work published 1985)
- Blackburn, S. (1998). Wittgenstein, Wright, Rorty and minimalism. *Mind*, 107(425), 157–181.
- Boghossian, P. (1989). The rule-following considerations. *Mind*, 98(392), 507–549.
- Boghossian, P. (2002). The rule-following considerations. In Miller & Wright (Eds.), *Rule-following and meaning* (pp. 141–187). Acumen.
- Brink, D. (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Burge, T. (2003). Perceptual entitlement. *Philosophy and Phenomenological Research*, 67(3), 503–548.
- Byrne, D. (2005). Compositionality and the manifestation challenge. *Synthese*, 144, 101–136.
- Chomsky, N. (1986). *Knowledge of language*. Praeger.
- Cornfield, D. (2003). *Towards a philosophy of real mathematics*. Cambridge, UK: Cambridge University Press.
- Davidson, D. (2001). A coherence theory of truth and knowledge. In *Subjective, intersubjective, objective* (pp. 137–153). Oxford: Oxford University Press. (Original work

- published 1983)
- Devitt, M. (1991a). Aberrations of the realism debate. *Philosophical Studies*, 61, 43–63.
- Devitt, M. (1991b). *Realism and truth* (2nd ed.). Blackwell.
- Dummett, M. (1959). Truth. *Proceedings of the Aristotelian Society*, 59, 141–162.
- Dummett, M. (1969). The reality of the past. *Proceedings of the Aristotelian Society*, 69, 239–258.
- Dummett, M. (1976). What is a theory of meaning? (II). In G. Evans & J. McDowell (Eds.), *Truth and meaning* (pp. 67–136). Oxford University Press.
- Dummett, M. (1978a). The philosophical basis of intuitionistic logic. In *Truth and other enigmas* (pp. 215–247). Duckworth. (Original work published 1973)
- Dummett, M. (1978b). Realism. In *Truth and other enigmas* (pp. 145–165). Duckworth. (Original work published 1963)
- Edwards, J. (1992). Best opinion and intentional states. *The Philosophical Quarterly*, 42(166), 21–33.
- Fodor, J. A. (1992a). A theory of content ii: The theory. In *A theory of content and other essays* (pp. 89–136). Mass.: MIT Press.
- Fodor, J. A. (1992b). A theory of content i: The problem. In *A theory of content and other essays* (pp. 51–88). Mass.: MIT Press.
- Geach, P. T. (1960). Ascriptivism. *Philosophical Review*, 69, 221–225.
- Geach, P. T. (1965). Assertion. *Philosophical Review*, 74, 449–465.
- Gibbard, A. (1990). *Wise choices, apt feelings*. Oxford: Clarendon Press.
- Hale, B. (1986). The compleat projectivist. *The Philosophical Quarterly*, 36(142), 65–84.
- Hale, B. (1997). Realism and its oppositions. In B. Hale & C. Wright (Eds.), *A companion to the philosophy of language* (pp. 271–308). Blackwell.
- Hume, D. (1969). *A treatise of human nature*. London: Penguin. (Original work published 1739)
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.
- Jackson, F., & Pettit, P. (1990). Program explanation: A general perspective. *Analysis*, 107–117.
- Joyce, R. (2001). *The myth of morality*. Cambridge: Cambridge University Press.
- Kim, J. (1993). Supervenience as a philosophical concept. In *Supervenience and mind* (pp. 131–160). Cambridge University Press. (Original work published 1990)
- Kripke, S. A. (1982). *Wittgenstein on rules and private language*. Oxford: Blackwell.
- Mackie, J. L. (1977). *Ethics: inventing right and wrong*. Penguin.
- McDowell, J. (1986). Critical notice: Ethics and the limits of philosophy. *Mind*, 95(379), 377–386.
- McDowell, J. (1996). *Mind and world*. Cambridge, Mass.: Harvard University Press. (Original work published 1994)
- McDowell, J. (1998a). Aesthetic value, objectivity, and the fabric of the world. In *Mind, value and reality* (pp. 112–130). Cambridge, Mass.: Harvard University Press. (Original work published 1983)
- McDowell, J. (1998b). Anti-realism and the epistemology of understanding. In *Meaning, knowledge, and reality* (pp. 314–343). Cambridge, Mass.: Harvard University Press. (Original work published 1981)
- McDowell, J. (1998c). Are moral requirements hypothetical imperatives? In *Mind, value and reality* (pp. 77–94). Cambridge, Mass.: Harvard University Press. (Original work published 1978)

- McDowell, J. (1998d). Criteria, defeasibility, and knowledge. In *Meaning, knowledge, and reality* (pp. 369–394). Cambridge, Mass.: Harvard University Press. (Original work published 1982)
- McDowell, J. (1998e). In defence of modesty. In *Meaning, knowledge, and reality* (pp. 87–107). Cambridge, Mass.: Harvard University Press. (Original work published 1987)
- McDowell, J. (1998f). Intentionality and interiority in Wittgenstein. In *Mind, value and reality* (pp. 297–321). Cambridge, Mass.: Harvard University Press. (Original work published 1991)
- McDowell, J. (1998g). Mathematical platonism and Dummettian anti-realism. In *Meaning, knowledge, and reality* (pp. 344–365). Cambridge, Mass.: Harvard University Press. (Original work published 1989)
- McDowell, J. (1998h). Meaning and intentionality in Wittgenstein’s later philosophy. In *Mind, value and reality* (pp. 263–278). Cambridge, Mass.: Harvard University Press. (Original work published 1993)
- McDowell, J. (1998i). Non-cognitivism and rule-following. In *Mind, value and reality* (pp. 198–218). Cambridge, Mass.: Harvard University Press. (Original work published 1981)
- McDowell, J. (1998j). One strand in the private language argument. In *Mind, value and reality* (pp. 279–296). Cambridge, Mass.: Harvard University Press. (Original work published 1989)
- McDowell, J. (1998k). On “the reality of the past”. In *Meaning, knowledge, and reality* (pp. 295–313). Cambridge, Mass.: Harvard University Press. (Original work published 1978)
- McDowell, J. (1998l). Projection and truth in ethics. In *Mind, value and reality* (pp. 151–166). Cambridge, Mass.: Harvard University Press. (Original work published 1987)
- McDowell, J. (1998m). Values and secondary qualities. In *Mind, value and reality* (pp. 131–150). Cambridge, Mass.: Harvard University Press. (Original work published 1985)
- McDowell, J. (1998n). Virtue and reason. In *Mind, value and reality* (pp. 50–73). Cambridge, Mass.: Harvard University Press. (Original work published 1979)
- McDowell, J. (1998o). Wittgenstein on following a rule. In *Mind, value and reality* (pp. 221–262). Cambridge, Mass.: Harvard University Press. (Original work published 1984)
- McDowell, J. (2000). Response to Crispin Wright. In C. M. Crispin Wright Barry C. Smith (Ed.), *Knowing our own minds* (pp. 47–62). Oxford University Press.
- McDowell, J. (2009a). Are meaning, understanding, etc., definite states? In *The engaged intellect* (pp. 79–95). Harvard University Press.
- McDowell, J. (2009b). Eudaimonism and realism in Aristotle’s ethics. In *The engaged intellect* (pp. 23–40). Harvard University Press. (Original work published 1995)
- McDowell, J. (2009c). Experiencing the world. In *The engaged intellect* (pp. 243–256). Harvard University Press. (Original work published 2000)
- McDowell, J. (2009d). How not to read *Philosophical Investigations*: Brandom’s Wittgenstein. In *The engaged intellect* (pp. 96–111). Harvard University Press. (Original work published 2002)
- McDowell, J. (2009e). Wittgensteinian “quietism”. *Common Knowledge*, 15(3), 365–372.
- McDowell, J. (2011). *Perception as a capacity for knowledge*. Marquette University Press.

- McGinn, C. (1984). *Wittgenstein on meaning*. Basil Blackwell.
- Miller, A. (2002). What is the manifestation argument? *Pacific Philosophical Quarterly*, 352–383.
- Miller, A. (2003a). *An introduction to contemporary metaethics*. Polity.
- Miller, A. (2003b). The significance of semantic realism. *Synthese*, 136, 191–217.
- Miller, A. (2003c). What is the acquisition argument? In A. Barber (Ed.), *Epistemology of language* (pp. 459–495). Oxford University Press.
- Miller, A. (2006). Realism, and anti-realism. In E. Lepore & B. C. Smith (Eds.), *Philosophy of language* (pp. 983–1005). Oxford University Press.
- Miller, A. (2007). Another objection to Wright's treatment of intention. *Analysis*, 67, 257–263.
- Miller, A. (2009). Primary qualities, secondary qualities and the truth about intention. *Synthese*, 171, 433–442.
- Miller, A. (2010). Non-cognitivism. In J. Skorupski (Ed.), *The Routledge companion to ethics* (pp. 321–333). London: Routledge.
- Miller, A. (2013). *An introduction to contemporary metaethics* (2nd ed.). Cambridge: Polity.
- Moore, G. E. (1903). *Principia ethica*. Cambridge: Cambridge University Press.
- Moore, G. E. (1921). *Ethics*. London: Williams and Norgate.
- Nida-Rmelin, M. (2010). Qualia: The knowledge argument. *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2010/entries/qualia-knowledge/>.
- Putnam, H. (1978). *Meaning and the moral sciences*. London: Routledge and Kegan Paul.
- Railton, P. (1986). Moral realism. *Philosophical Review*, 95, 163–207.
- Sellars, W. (1997). *Empiricism and the philosophy of mind*. Harvard University Press. (Original work published 1956)
- Shieh, S. (1998). Undecidability in anti-realism. *Philosophia Mathematica*, 6, 324–333.
- Smith, M. (1994). *The moral problem*. Oxford: Blackwell.
- Thurston, W. P. (2006). On proof and progress in mathematics. In R. Hersh (Ed.), *18 unconventional essays on the nature of mathematics* (pp. 37–55). Springer. (Original work published 1994)
- Weiss, B. (2002). *Michael Dummett*. Chesham, England: Acumen.
- Williams, B. (1978). *Descartes: The project of pure enquiry*. London: Penguin Books.
- Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Kegan Paul, Trench, Trubner & Co., Ltd.
- Wittgenstein, L. (1978). *Remarks on the foundations of mathematics*. Oxford: Basil Blackwell. (Original work published 1956)
- Wittgenstein, L. (2001). *Philosophical investigations*. Oxford: Blackwell. (Original work published 1953)
- Wright, C. (1980). *Wittgenstein on the foundations of mathematics*. Cambridge: Harvard University Press.
- Wright, C. (1982). Strict finitism. *Synthese*, 51, 203–282.
- Wright, C. (1984). Kripke's account of the argument against private language. *Journal of Philosophy*, 81(12), 759–778.
- Wright, C. (1988). Moral values, projection and secondary qualities. *Proceedings of the Aristotelian Society, Supplementary Volume*, 1–26.

- Wright, C. (1989a). Wittgenstein's later philosophy of mind: Sensation, privacy, and intention. *The Journal of Philosophy*, 86(11), 622–634.
- Wright, C. (1989b). Wittgenstein's rule-following considerations and the central project of theoretical linguistics. In A. George (Ed.), *Reflections on Chomsky* (pp. 233–264). Blackwell.
- Wright, C. (1992). *Truth and objectivity*. Cambridge, Mass.: Harvard University Press.
- Wright, C. (1993a). Introduction. In *Realism, meaning and truth* (2nd ed., pp. 1–43). Blackwell.
- Wright, C. (1993b). Misconstruals made manifest. In *Realism, meaning and truth* (2nd ed., pp. 239–261). Blackwell. (Original work published 1989)
- Wright, C. (1993c). Realism, truth-value links, other minds and the past. In *Realism, meaning and truth* (2nd ed., pp. 85–106). Blackwell. (Original work published 1980)
- Wright, C. (2000). Self-knowledge: The Wittgensteinian legacy. In C. M. Crispin Wright Barry C. Smith (Ed.), *Knowing our own minds* (pp. 13–45). Oxford University Press. (Original work published 1998)
- Wright, C. (2007). Rule-following without reasons: Wittgenstein's quietism and the constitutive question. *Ratio*, 20(4), 481–502.